

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 2, 2003, 11:19:48 ; Search time 5038 Seconds  
(without alignments)  
3386.128 Million cell updates/sec

Title: US-09-469-200D-2

Perfect score: 2145  
Sequence: 1 MRTLKNIITVAFSIFWLL.....KLYSLFTIRNADWGTREKLL 417

Scoring table:  
BLOSUM62 Xgapop 10.0, Xgapext 0.5  
Xgapop 10.0, Ygapext 0.5  
Ygapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09469200/runat\_02102003\_090348\_22777/app\_query.fasta.1.583  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09469200.ecgn\_1\_13508.ernat\_02102003\_090348\_22777 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPOP=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rod.\*  
36: em.htg\_nam.\*  
37: em.htg\_vrt.\*  
38: em.sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2145	100.0	1254	1	AF023876 Streptoco
2	2145	100.0	1254	6	BD087249 Hyaluroa
3	2145	100.0	3552	1	AV173078 Streptoco
4	2138	99.7	7268	1	AF347022 Streptoco
5	1695.5	79.0	3466	1	SUB242946
6	1588.5	74.1	1350	1	STRHASAO
7	1588.5	74.1	10851	1	AE010123
8	1586.5	74.0	1440	1	STRHASA
9	1586.5	74.0	1512	6	AR232201
10	1586.5	74.0	9844	1	AE006637 Streptoco
11	1585.5	73.9	53896	1	AE014172 Streptoco
12	1585.5	73.9	323825	1	AF005146 Streptoco
13	1052.5	49.1	181654	1	AF065404 Bacillus
14	1052.5	49.1	181677	1	AE011190
15	745	34.7	1893	12	AF518732
16	502	23.4	6544	1	AF284858 Bradythiz
17	493.5	23.0	110000	1	REU80928_0
18	484.5	22.6	10973	1	AF522456 Sinorhizo
19	479	22.3	346820	1	AP003008 Mesorhizo
20	477	22.2	16124	1	RSU53327 Rhizobium s
21	476.5	22.2	6014	1	RGNOG
22	476	22.2	152050	1	ML0672113 Mesorhizo
23	475	22.1	3373	1	RNODABC
24	475	22.1	3403	1	RTNODABCD
25	474	22.1	3304	1	AF038577
26	474	22.1	4230	1	SYLNOD
27	474	22.1	14188	1	AE007237 Sinorhizo
28	472	22.0	33038	1	AF222753 Bradythiz
29	464	21.6	1388	1	RLNODC
30	464	21.6	3094	1	RHNOODABC
31	463	21.6	10062	1	AE000076 Rhizobium
32	463	21.6	320040	6	A79350
33	463	21.6	320040	6	A93002
34	460	21.4	3080	1	RLNOD2
35	454.5	21.2	2627	1	RLNODABC
36	454.5	21.2	12055	1	RLNOD
37	451	21.0	4317	1	ACNODI
38	451	21.0	9549	1	RHANODABC
39	447.5	20.9	1994	9	AF234839
40	447.5	20.9	4049	9	AF232772 Homo sapi
41	442	20.6	2465	10	AB097569 Rattus no
42	441.5	20.6	2370	5	AF190742 Danio rer
43	440	20.5	1659	4	AB055979 Oryctolag
44	440	20.5	1665	6	E30973 Hyaluronic
45	440	20.5	5919	10	MMU86408 Mus musculu

ALIGNMENTS

RESULT 1

RESULT 2				
BD087249	LOCUS	1254 bp	DNA linear	PAT 27-AUG-2002
	DEFINITION	Hyaluronan synthase gene and utilization thereof.		
	ACCESSION	BD087249		

```

VERSION      BD087249.1 GI:22632859
KEYWORDS
SOURCE       JP 2001521741-A/1.
ORGANISM     Streptococcus dysgalactiae subsp. equisimilis
              Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
              Streptococcus.
REFERENCE    1 (bases 1 to 1254)
AUTHORS     Weigel,P.H., Kumari,K. and Deangelis,P.
TITLE       Hyaluronan synthase gene and utilization thereof
JOURNAL     Patent: JP 2001521741-A 1 13-NOV-2001;
              THE BOARD OF REGENTS OF THE UNIVERSITY OF OKLAHOMA
COMMENT     OS Streptococcus equisimilis
              PN JP 2001521741-A/1
              PD 13-NOV-2001
              PF 30-OCT-1998 JP 2000519083
              PR 31-OCT-1997 US 60/064435, 26-OCT-1998 US 09/178851 P1
              PC PAUL H WEIGEL,KSHAMA KUMARI,PAUL DEANGELIS
              C12N15/09,A61K31/728,A61P43/00,C12N1/21,C12N9/10,C12P19/04, PC
              C12Q1/68//
              PC (C12N15/09,C12R1:46),C12N15/00,C12R1:46) CC
              Hyaluronan synthase gene and utilization thereof FH Key
FEATURES     Location/Qualifiers
              FT source 1..1254
              /organism="Streptococcus equisimilis".
              /mol_type="genomic DNA"
              /db_xref="taxon:119602"
              335 a 240 c 261 g 418 t
BASE COUNT  335 a 240 c 261 g 418 t
ORIGIN
Alignment Scores:
Pred. No.: 8.47e-183 Length: 1254
Score: 2145.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-469-200D-2 (1-417) x BD087249 (1-1254)
QY 1 MetArgThrLeuLysAsnLeuIleThrValValAlaPheSerIlePheTrpValLeuLeu 20
Db 1 ATGAGAACATTAAAACCTCATAACTGCTGGCCCTTAGTATTATTTTGGGTACTGTTG 60
QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40
Db 61 ATTACGTCATAGTTTATCTCTTGGTGCTAAAGGAAGCTTGTCAATTTATGGCTTTTG 120
QY 41 LeuIleAlaTyrLeuLeuValLysMetSerLeuSerPheTyrLysPropheLysGly 60
Db 121 CTGATAGCTACCTATTAGTCAAAATGCTTATCTTTTACAGCCATTATAAGGA 180
QY 61 ArgAlaGlyGlnTyrLysValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSer 80
Db 181 AGGCTGGCAATTAAGTTGACGCAATTAATCCCTCTATAAGCAAGATGCTGAGTCA 240
QY 81 LeuLeuGluThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVal 100
Db 241 TTGCTAGAGACCTTAAAAGTGTTCAGCAGCAACCTTATCCCTAGCAAAATTTATGTT 300
QY 101 ValAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAsp 120
Db 301 GTTGACGATGGAAGTGTGATGAGACAGGATTAAGCCATTAAGACATATGCGGTGAC 360
QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140
Db 361 ACTGGTGACCTATCAAGCAATGTCAATTTTCATCGGTTCAGAGAAAATCAAGGAAGCGT 420
QY 141 HisAlaGlnAlaTrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160

```

---

```

Db 421 CATGCACAGGCTGGCCCTTTGAAGATCATCAGCGCTGATGCTCTTTTACGCGTTGACTCA 480
QY 161 AspThrTyrIleTyrProAspAlaLeuGluGluLeuLysThrPheAsnAspProThr 180
Db 481 GATACATTATATCTACCTGATGCTTTAGAGAGGTTGTTAAAACCTTTAATGACCCAACT 540
QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200
Db 541 GTTTTGTGCGACGGGTCCCTTAATGTCAGAAATAGACAAACCAATCTCTTACACGC 600
QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThr 220
Db 601 TTGACAGATATTCGCTATGATTAATGCTTTTGGCGTTGAACGAGCTGCCCAATCCGTTACA 660
QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValValPro 240
Db 661 GGTAATATCTTTGTTGCTCAGGTCGCTTAGCGTTTACAGCGGAGGTGGTTGTTCT 720
QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260
Db 721 AACATAGATAGATACATCAACAGACCTTCTGGGTATTCCTGTAAGTATGGTGATGAC 780
QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLys 280
Db 781 AGGTGCTTGACCAACTATGCAACTGATTTAGGAAGACTGTTTATCAATCCACTGCTAAA 840
QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsn 300
Db 841 TGTATTACAGATGTTCTCGACAGATGCTCTACTTGAAGCGCAAAACCGTGGAC 900
QY 301 LysSerPhePheArgGluSerIleIleSerValLysLysIleMetAsnAsnPropheVal 320
Db 901 AAGTCCCTCTTTAGAGAGTCCATATTCTGTTAAGAAATCATGAACATCCCTTTGTA 960
QY 321 AlaLeuTrpThrIleLeuGluValSerMetPheMetIleValTyrSerValValAsp 340
Db 961 GCCCTATGACCATCTATTCAGGTGCTATGTTTATGATGCTTGTATTCTGGTGGAT 1020
QY 341 PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle 360
Db 1021 TTCCTTTAGGCANTGTCAGAGATTTGATTGGCTCAGGGTTTAGCCTTCTGCTGATT 1080
QY 361 IlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPhe 380
Db 1081 ATCTTCATCTTGCCCTGTGCGAACATTCATTCATGCTTAAGCACCCGCTGCTCTTC 1140
QY 381 LeuLeuSerPropheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyr 400
Db 1141 TTGTTATCTCCGTTTATGGGGTGTGCATTTGTTTGTCTACAGCCCTTGAATATATAT 1200
QY 401 SerLeuPheThrIleArgAsnAlaAspTrpGlyThrArgLysLysLeuLeu 417
Db 1201 TCTCTTTTACTATTAGAAATGCTGACTGGGACACGTAATAAAATATTA 1251
RESULT 3
AY173078 3552 bp DNA linear BCT 15-JAN-2003
LOCUS Streptococcus equi subsp. zooepidemicus hyaluronid acid operon,
DEFINITION partial sequence.
VERSION AY173078.1 GI:27762599
KEYWORDS
SOURCE Streptococcus equi subsp. zooepidemicus
ORGANISM Streptococcus equi subsp. zooepidemicus
          Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
          Streptococcus.
REFERENCE 1 (bases 1 to 3552)
AUTHORS Wu,X. and Gao,H.
TITLE Molecular characterization of hyaluronid acid operon from group C
JOURNAL Streptococcus equi subsp. zooepidemicus
          Unpublished
REFERENCE 2 (bases 1 to 3552)
AUTHORS Wu,X., Gao,H. and Chen,J.

```

**TITLE** Molecular cloning, expression, and characterization of hasC and galU, two genes encoding UDP-glucose pyrophosphorylase in group C Streptococcus equi subsp. zooepidemicus

**JOURNAL** Unpublished

**REFERENCE** 3 (bases 1 to 3552)

**AUTHORS** Wu, X. and Gao, H.

**TITLE** Direct Submission

**JOURNAL** Submitted (05-NOV-2002) Department of Biology, Tsinghua University, Beijing 100084, China

**FEATURES**

**source**

1..3552

Location/Qualifiers

/organism="Streptococcus equi subsp. zooepidemicus"

/mol\_type="genomic DNA"

/sub\_species="zooepidemicus"

/db\_xref="taxon:40041"

1..3552

/gene="hyaluronidase"

1..1254

/gene="hasA"

1..1254

/gene="hasB"

/note="required for hyaluronidase synthesis"

/codon\_start=1

/transl\_table=11

/product="hyaluronan synthase"

/protein\_id="AAO20097.1"

/db\_xref="GI:27762600"

/translation="MTLKNLTVAFSIFWVLLIYVNVYLFAGKSGLSIYGLLIAYLVKMSGLTFKPKAGQYKVAALIFSYNEDAESLLEKASVQCCQYPLAELIYVDDGADATGKRIEDYVRDQGLSSNVVHRSKRNQKRAQAWAFERSDADVLTVDYDPALELLATFNFTFAATGHLNVRNQTLLRLDIDYDNFAGVERAAQSVTGILVSGSLSVREYVFNIDRYNOTFLGIPVSGIDRCILNTYATDLGKTYIQSTAKCTDVPKMGTYLQKQNRWKSFFRESISVKIMNPFVALWTILEVSNFVMLVSVDFVGVNRFEDLRLVLAFLVLIIFVALCRNIHYMLKPLSFLSLPPYGVHLFVLOPLKSLSTIRNADWGRKLL"

1507..2712

/gene="hasB"

1507..2712

/gene="hasB"

/codon\_start=1

/transl\_table=11

/product="UDP-glucose dehydrogenase"

/protein\_id="AAO20098.1"

/db\_xref="GI:27762601"

/translation="MKISVAGSYVGLSILLAQHNDVTVVDIIDEKVLINGLISPKDADTEYLKAPLNTATLDGASAYSNADLIITAPTNIDSEYVDFRHYEVIEQVLDNASTIILIKSTIPGFIKHREKYQDRIIFSPFELRESKALYDNLVPSRLIVSYKDSRPIVQAKAKAFAGLLKEGAKSDTPVLPWGSQEAQVLEFANFLANRYSYFNELDYSKGLDQQRVIEGVHQDRIGNHNNPSPFGYICLPKXSKQLLIANTRGIPQSLMSAIVESNKRISYLAQIILDRASSQKQGVPLTIGFYRLIMKSNDRFRESAIDIIDINDGVNVIYEFMLGEDIGYRVVVKDLEQFKNESTIIVSNRFEDDLGDVLDKYTRDVFGRD"

2762..3552

/gene="hasC"

2762..3552

/gene="hasC"

/codon\_start=1

/transl\_table=11

/product="UDP-glucose pyrophosphorylase"

/protein\_id="AAO20099.1"

/db\_xref="GI:27762602"

/translation="MTKVRKAIIPAAAGLRPLPATKALAKEMLPVVKPTIQIIVEEALKAAGTEILVYTKAKRSIEDHDFNFEINLQAKGTELLKLVDEITAINLHFTIRQSHPRGLGDAVLQAKAFVGNPFPVLMGDDLDITNPISAKPLTKLIEDYDCTHASTIIVMVRPHEVSNYGVIAPOGKAVKGLYSVTEVEKPSDEAPSDLAIGRVLLRPEIFAILLENQAPGAGNEVOLITDAIDKLNKTQRYFAREFKGRYDVGDKFG"

**BASE COUNT** 1006 a 732 c 817 g 997 t

**ORIGIN**

Alignment Scores:

Pred. No.: 3,8e-182 Length: 3552

Score: 2145.00 Matches: 417

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 1 Gaps: 0

US-09-469-200d-2 (1-417) x AY173078 (1-3552)

QY 1 MetAqrThrLeuLysAsnLeuLeuThrValValAlaPheSerIlePheTrpValLeuLeu 20

DB 1 ATGAGAACATTAAAAACCTCATAACTCTGTGGCCTTTAGTATTTTGGGTACTGTG 60

QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40

DB 61 ATTACGTCATGTTATCTCTTTTGGTCTAAAGAGAGCTTCTCAATTTATGCTTTTG 120

QY 41 LeuIleAlaTyrLeuLeuValLysMetSerLeuSerPhePheTyrLysPropheLysGly 60

DB 121 CTGATAGCTTACCTTATAGTCAAAATGCTTATCTCTTTTCAAGCCATTTAAGGGA 180

QY 61 ArgAlaGlyGlnTyrLysValAlaAlaIlePheProSerTyrAsnGluAspAlaGluSer 80

DB 181 AGGCTGGGCAATATAGGTTGCACCCATTTATCCCTCTTATACGAGATGCTGAGTCA 240

QY 81 LeuLeuGlnThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVal 100

DB 241 TTGCTAGAGACCTTAAAAAGTGTTCAGCAGCAAAACCTATCCCTAGCAGAAATTTATGTT 300

QY 101 ValAspAspGlySerAlaAspGlnThrGlyIleLysArgIleGluAspTyrValArgAsp 120

DB 301 GTTGAGATGGAAGTGTGATGAGCAGGATTAAGCGCTTGAAGACTATGTCGCTGAC 360

QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140

DB 361 ACTGTTGACCTATCAAGCAATGTCATTGTCATCGGTTCAGAGAAAAATCAAGAAAGCGT 420

QY 141 HisAlaGlnAlaTyrAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160

DB 421 CATGCACAGGGCTGGGCTTTGAAAGATCAGACGCTGATGCTTTTTCACCGTTGACTCA 480

QY 161 AspThrTyrIleTyrProAspAlaLeuGluLeuLysThrPheAsnAspProThr 180

DB 481 GATCTTATATCTACCTGATGCTTTAGAGAGTGTGTTAAACCTTTATGACCAACT 540

QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200

DB 541 GTTTTTCCTGGCAGCGGTCACTTAATGTCAGAAATAGACAAACCAATCTCTTAACACGC 600

QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThr 220

DB 601 TTGACAGATTTTCGCTATGATATGCTTTTGGCGTTGACGAGCTGCCCAATCCGTTTACA 660

QY 221 GlyAsnIleLeuValLysSerGlyProLeuSerValTyrArgGluValValPro 240

DB 661 GGTAATATCTTGTGCTCAGGTCGCTTAGCGTTTACAGACGCGAGGTGTTGTCCT 720

QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260

DB 721 AACATAGATAGATACATCAACACGACCTTCTCTGCTGATTAAGTATGTTGCTGATGAC 780

QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLys 280

DB 781 AGGTGCTTGACCACTATGCAACTGATTTAGGAAGAGCTGTTTATCAATCCATCGCTAA 840

QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsn 300

DB 841 TGTATTACAGATGTTCTCTGACAGAGTGTCTACTTACTTGAAGCAGCAAAACCGCTGGAAC 900

QY 301 LysSerPheArgGluSerIleIleSerValLysIleMetAsnAsnPropheVal 320

DB 901 AAGTCCTTCTTTAGAGATGCTTATTTCTGTTAAGAAATCATGACCAATCTCTTTGTA 960

QY 321 AlaLeuThrThrIleLeuGluValSerMetPheMetMetLeuValTyrSerValValAsp 340

DB 961 GCCCTATGAGACCACTACTTGAGGTGCTATGTTTATGATGCTTGTATTCTGCTGCTGGAT 1020

```

QY 341 PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle 360
Db 1021 TTTTGTAGGCAATGTCAGAGAAATTCATTTGGCTCAGGGTTTAGCTTCTCGGTGAT 1080
QY 361 IlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPhe 380
Db 1081 ATCTTCATTTGTCCTGTGTGGACATTCATTACATGCTTAAGCACCGCTGCTCTTC 1140
QY 381 LeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyr 400
Db 1141 TTCTTATCTCGCTTTATGGGTGCTGCATTTGTTGCTCTACAGCCCTTGAATATATAT 1200
QY 401 SerLeuPheThrIleArgAsnAlaAspTrpGlyThrArgLysLysLeuLeu 417
Db 1201 TCTCTTTTACTATAGAAATGCTGACGGGACACGTAATAAAATTAATTA 1251

RESULT 4
AF347022 7268 bp DNA linear BCT 13-NOV-2002
LOCUS Streptococcus equi subsp. zooepidemicus has operon, partial
DEFINITION
ACCESSION AF347022
VERSION AF347022.1 GI:24940621
KEYWORDS
SOURCE Streptococcus equi subsp. zooepidemicus
ORGANISM Streptococcus equi subsp. zooepidemicus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 7268)
Blank,L.M., Hugenholtz,P. and Nielsen,L.K.
Cloning and characterization of the hyaluronic acid synthesis (has)
operon from Streptococcus equi subsp. zooepidemicus
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 7268)
Blank,L.M. and Nielsen,L.K.
Direct Submission
TITLE Submitted (09-FEB-2001) Chemical Engineering Department, The
University of Queensland, Coopers Road, Building #74, Brisbane,
Queensland 4072, Australia
FEATURES
source
1..7268
/organism="Streptococcus equi subsp. zooepidemicus"
/mol_type="genomic DNA"
/sub_species="zooepidemicus"
/db_xref="taxon:40041"
469..>7268
/gene="has operon"
-10_signal 469..474
/gene="has operon"
mRNA 481..>7268
/gene="has operon"
/evidence=experimental
512..1784
/gene="has"
RBS 512..517
/gene="has"
CDS 531..1784
/gene="has"
/codon_start=1
/transl_table=11
/product="hyaluronic acid synthase"
/protein_id="AAN65248.1"
/db_xref="GI:24940622"
/translators="MSTLKNLTVVAFSIFWLLIYVNVYLFAGKSLSYGFLLIAY
LVKMSLFFYKFKRAGQYKVAALIPSDNEDASLETLKSVQQQYPLAEITVD
DSADETGIKRIEDYVRDGLSSNVIVHSEKNQKRAQAFERSADVFLTVD
DTIYPDALIELLKTNDPTFAATGHLNVRNQTNLITRDIYDNAGFVERAQS
VTGNLVCSGPLSVYREVVVNPIDRYINQTFILGIPVSGIDRCLTNYATDLGKTVYQ
STAKCITDPDKMSTYLVKQONKNSFFRESITSVKKIMNPFVALTILEYSMMML
VYSVDYFVDNREFDMLVLAFLVLIETVALCRNIHYMKRPLSFLISFFYGLHLF
VLOPLKYSIPTIRNADWGTAKLL"
2023..3242
gene

```

```

/gene="hasp"
2023..2028
/gene="hasp"
2037..3242
/gene="hasp"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="UDP-glucose dehydrogenase"
/protein_id="AAN65249.1"
/db_xref="GI:24940623"
/translators="MKISVAGSGYGLSILLQAHDVVVDIIDEKVRILNQGISP
IKDADTEFLKNAPLNTATDGCASAYSNADLIITPTNYLSEKALYDNLVPSRLIV
OVLIDNASATIIKSTIPGLFKHREKYQDRILISPELSEKALYDNLVPSRLIV
SEKDSRPTQAARAFAGLLKAGAKSDTPTVLFMSQAEAYKLFANTFLANRYSYF
NEUDTYSKGLDQARVIEGVCHDQRIQHNHNFSGYGLCPKDSKQLLANRGYIP
QSLMSAIVSEKNIRKSYLAQIILDRASSQKQAGPLTIGTYILNKSNSDNFRESAIP
DIIDIINDYGVNIVIEPEMLGIDGYRVVKOLEQFKNESTIIVSNRPEDDIDGVIDKY
YTRDVFGRD"
3280..4194
/gene="hasc"
3280..3285
/gene="hasc"
3292..4194
/gene="hasc"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="UDP-glucose pyrophosphorylase"
/protein_id="AAN65250.1"
/db_xref="GI:24940624"
/translators="MTKVKRAIIPAGLGTFLPATKALAKEMLPVDKPTIOPIVEE
ALKAGEELTVVTKAKRSIEDHDSNFELEYNLOAKGTELLKLVDETLNHLSTR
QSHPRGLDGYLOAKAFYGVNEPVMGLDMDITNPSAKPLAKLIIEDICTHASTII
AVMRPHEEVSNIQYAPQKAGVGLSVETFEKPSDFADSLAIIGRTLLTPEIF
ALLENQAPAGNEQLADAIDKLNKTRVRFAREFKRGVDVGVGTFKMTSLDYALKH
POVKDLDTDYIIKLSKQINKDKVK"
4349..5744
/gene="glmU"
4349..4354
/gene="glmU"
4362..5744
/gene="glmU"
/codon_start=1
/transl_table=11
/product="UDP-N-acetyl-glucosamine pyrophosphorylase"
/protein_id="AAN65251.1"
/db_xref="GI:24940625"
/translators="MKNYAIILAAAGKGTMRNSGLSKVLHKVSGLSMLSEVLKSVSALA
POKLTAVIGHQAEQVRAVLGDQLTIVQEEQLGTGHAVMAEEELSGLEGTLVIAGD
PTLIRGESLKALDYHIREKNVATILTNADKDPFGYGRIRNAGEVYVNIVEOKDANE
AROEVEKINTGYIFDNKRLFEALHLLTDNAQGEYVLTDLISIFKASQEKVAYLLK
DFDESGLVNDLALAQAQAEVIMQERINKQHLNGVTLQNPATYIIESSVEIAPVILBA
NVILKQTRIGRSYNTSIILDSRIGEGVVVSQSVIGSVLADGVTVGTFIAHIRPD
SQDCECHVHIGNEVEVKGHIGANTKAGHTYLGNAEIGSEVNIAGSITVNDGQRKY
QTVIGDHAFIGSHSTLIAPVEVGENALTAAGSTIAQSVPADSVAIGRSRQVKEGYAK
RLPHHPDQPK"
5905..7268
/gene="pgi"
5905..5910
/gene="pgi"
5919..7268
/gene="pgi"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate isomerase"
/protein_id="AAN65252.1"
/db_xref="GI:24940626"
/translators="MSHTTFDYSKLEQFAGQHEIDFLQGVTEADQALROGTGPGSD
FLWGLLELPNDYKFEFARILKAAEKIKADSDVLVVICGIGSVYLGAKAIDFLNSHFAN
FLWAKERKAPQILLIAGNSISSVLAIDLVDYQDKDFSVNYSKSTTTERALAFYFK
ELIVKYGQEEANKRIYATTDKVKYKAVKEADANHWETFVVPDNPVGRFSSVITAVGLF
PIAASGADITALMEGANARKDLSSDKISENIATQYAVVRNLIYRKGVYTEILANYEP

```

SLOVESHWKOLAGESEKDKGIYPTSANESTDLHSLGQFIOEGYRNLFETVIRYDK  
 PRQNIIPMAEDLDGLGGLQGRDVFYNNKATDGLVLAHTDGVNFMFIIIPQDDEF  
 TLGTTIIFELIALSLGVLNGVNFDPQGVYAKNNMFALLKPGFELGAALNARI\*  
 BASE COUNT 2105 a 1485 c 1733 g 1945 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4,53e-181 Length: 7268  
 Score: 2138.00 Matches: 416  
 Percent Similarity: 99.76% Conservative: 0  
 Best Local Similarity: 99.76% Mismatches: 1  
 Query Match: 99.67% Indels: 0  
 DB: 1 Gaps: 0

US-09-469-200d-2 (1-417) x AF347022 (1-7268)

QY 1 MetArgThrLeuLysAsnLeuThrValValAlaPheSerIlePheTrpValLeuLeu 20  
 Db 531 ATGAGACATTAATAAACCTCATACCTGTGTGGCCCTTAGTATTTTGGGTACTGTG 590  
 QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40  
 Db 591 ATTTACGTCATATTTATCTCTTGGTCTAAGGAAGCTTGTCAATTTATGCTTTTGTG 650  
 QY 41 LeuIleAlaTyrLeuLeuValLysMetSerLeuSerPhePheTyrLysProPheLysGly 60  
 Db 651 CTGATAGCTTACCTATTAGTCAAAATGTCCTTATCCTTTTTCACAGCCATTAAAGGGA 710  
 QY 61 ArgAlaGlyGlnTyrLysValAlaAlaIleProSerTyrAsnGluAspAlaGluSer 80  
 Db 711 AGGCTGGGCATTAAGGTTCAGCCATTATTCCTCTTATAAGGAAGTGTGAGTCA 770  
 QY 81 LeuLeuGlnThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVal 100  
 Db 771 TTGCTAGAGCTTAAAGTGTTCAGCAGCAACCTATCCCTTAGCAGAAATTTATGTT 830  
 QY 101 ValAspAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAsp 120  
 Db 831 GTTGACATGGAAGTGTGATGAGACAGGTATTAAGCGATTGAAGACTATGTGCGTGAC 890  
 QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140  
 Db 891 ACTGTGACCTTACAGCAATGTCTATGTCACCGTCCAGAAATCAAGGAAGCGT 950  
 QY 141 HisAlaGlnAlaThrAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160  
 Db 951 CATGCACAGCGCTGGCGCTTGAAGATCAGACGCTGATGCTTTTGGACCGTCACTCA 1010  
 QY 161 AspThrTyrIleTyrProAspAlaLeuGluGluLeuLysThrPheAsnAspProThr 180  
 Db 1011 GATACTTATCTACCTGATGCTTTAGAGAGTGTCTTAAACCTTTAATGACCAACT 1070  
 QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200  
 Db 1071 GTTTTGTGCGAGGGTCACTTAATGTCAGAAATAGACAAACCATCTCTTACACGC 1130  
 QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaGlnSerValThr 220  
 Db 1131 TTGACAGATATTCGCTATGATTAATGCTTGGGTGTAAGAGAGTGTCCCAATCCGTTACA 1190  
 QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValValValPro 240  
 Db 1191 GGTATATCTCTGCTGCGAGCGCGCTTAGCGTTTACAGACGCGAGGTGTGTTGCT 1250  
 QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260  
 Db 1251 AACATAGATAGATACATCAACAGACCTTCTGGGTATTCCTGTAAGTATCGGTGATGAC 1310  
 QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyIleThrValTyrGlnSerThrAlaLys 280  
 Db 1311 AGGTGCTTGACCAACTGCAACTGATTTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 1370  
 QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsn 300

Db 1371 TGTATTACAGATGTTCTCTGCAAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAAC 1430  
 QY 301 LysSerPheArgGluSerIleIleSerValLysLysIleMetAsnAsnProPheVal 320  
 Db 1431 AAGTCCTTCTTAGAGAGTCCATTTCTGTTAAGAAATCATGAACAATCCCTTTGTA 1490  
 QY 321 AlaLeuThrTrpIleLeuGluValSerMetPheMetMetLeuValTyrSerValValAsp 340  
 Db 1491 GCCCTATGACCATACTTGAAGTCTCTATGTTTATGATGCTGTTTATCTGTTGGTGGAT 1550  
 QY 341 PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle 360  
 Db 1551 TTCTTTGTAGACAATGTCAGAAATTTGATTGGCTCAGGGTTTGGCCTTCTGGTGGATT 1610  
 QY 361 IlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPhe 380  
 Db 1611 ATCTTCATGTTGCTCTTGTGCTAATATTCATATATGCTTAAGCACCCTGCTGCTTC 1670  
 QY 381 LeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyr 400  
 Db 1671 TTGTTATCTCCGTTTATGGGGTACTGCAATTTGTTGCTACAGCCCTTGAATTTGAT 1730  
 QY 401 SerLeuPheThrIleArgAsnAlaAspTrpGlyThrArgLysLysLeuLeu 417  
 Db 1731 TCTCTTTTACTATTAGAAATGCTGACTGGGACACGATAAAATATTATA 1781

# RESULT 5

SUB242946 3466 bp DNA linear BCT 06-JAN-2001  
 LOCUS Streptococcus uberis hasA and hasB genes for putative hyaluronan  
 DEFINITION synthase and UDP-glucose dehydrogenase.  
 ACCESSION AJ242946  
 VERSION AJ242946.2 GI:7532762  
 KEYWORDS hasA gene; hasB gene; hyaluronan synthase; UDP-glucose  
 dehydrogenase.  
 SOURCE Streptococcus uberis  
 ORGANISM Streptococcus uberis  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1  
 AUTHORS Ward, P.N., Field, T.R., Ditcham, W.G., Maguin, E. and Leigh, J.A.  
 TITLE Identification and disruption of two discrete loci encoding  
 hyaluronic acid capsule biosynthesis genes hasA, hasB, and hasC in  
 Streptococcus uberis  
 JOURNAL Infect. Immun. 69 (1), 392-399 (2001)  
 MEDLINE 20569205  
 PUBMED 11119529  
 REFERENCE 2  
 AUTHORS Ward, P.N.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-JUN-1999) Ward P.N., Environmental Microbiology,  
 Institute for Animal Health, Compton Laboratory, Compton, nr  
 Newbury, Berkshire, RG20 7NN, UNITED KINGDOM  
 REVISION 3 (bases 1 to 3466)  
 REMARK Ward, P.N.  
 REFERENCE Direct Submission  
 AUTHORS Submitted (07-APR-2000) Ward P.N., Environmental Microbiology,  
 Institute for Animal Health, Compton Laboratory, Compton, nr  
 Newbury, Berkshire, RG20 7NN, UNITED KINGDOM  
 JOURNAL On Apr 10, 2000 this sequence version replaced gi:5441847.  
 COMMENT Location/Qualifiers  
 FEATURES  
 source  
 1..3466  
 /organism="Streptococcus uberis"  
 /mol\_type="genomic DNA"  
 /strain="0140J"  
 /db\_xref="taxon:1349"  
 /country="United Kingdom"  
 23..1276  
 /gene="hasA"  
 23..1276  
 /gene="hasA"

```

/function="putative biosynthesis of hyaluronic acid"
/codon_start=1
/transl_table=11
/product="hyaluronan synthase"
/protein_id="CAB46918.2"
/db_xref="GI:7532763"
/db_xref="SPTREMBL:Q9X9L5"
/translation="MEKLNLTIPWFIFLWLLIIGLVNFVFGTKSLTVYGIILLTY
LSIKMGLSFYRYPKSGVGQKVAALIPSYNEDGVLGLETLSKVQKQYTPAEIFVID
DGSVDKTKIKLVYEDYKLVNGFGQVIVHQPENPVKGRHAQAFERSADADFLVDSV
TYIYPDALBELLKTFNDPEVTAATGHNNRNQTNLLTRLDIRYDNAGFVERAAQSV
TGNLVCSPGLSYRRSVGIPNLEKVTSGTFLGVPVSGIDRCLTNATDLGKTVQVS
TARCDTDVDPKFKVFKQONRWKNSFEFESIIISVKLLATPSPVAVWTLEVSMEILV
YSFSLIGEAQENLKLAFVLIIFVVALCRNWHVYKHPFAFLSPFFGLHLFV
LQPLKYSFTURNATWETRKYSK"
1301..2506
/gene="hasB"
1301..2506
/gene="hasB"
/function="putative biosynthesis of D-glucuronic acid"
/codon_start=1
/transl_table=11
/product="putative UDP-glucose dehydrogenase"
/protein_id="CAB46919.2"
/db_xref="GI:7532764"
/db_xref="SPTREMBL:Q9X9L4"
/translation="MKIIVAGSGYVGLSVLLAKNPVTVDLIEKKVNLINQKQSP
IQVDYENILKSKLQRAVLADQAFROADLILITPTNDYDKEKFFDTSVETVIE
KALANSQALLVKTIPLGFKMKQKQYOTDRIIFSPFERESKALKDNLVPSKRIIV
SFEDDSNEVIEAKTFAQLKKGSLDKQVPLFMGSAEAEVAKLFANTYLAIRVYF
NELTYAKENGRVDNIIEGVCHDRIGIHYNNPSPFGYGYCPKDKTKOLLAGVDGP
QSLKATVDSNKRKEYTASQIQLQSLSDINVDPKDATIGYRLIMKSNDSNFRSAIK
DIIDHISKYQINLVYEPNMEDPDLPIIDLSDFKAMSHIIVSNRYDLALEDYKRV
YTRDIYGVDP"
complement(2703..>3466)
/feature="ORF1"
/codon_start=3
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB87142.1"
/db_xref="GI:7532765"
/db_xref="SPTREMBL:Q9L4H1"
/translation="QAFSVLYLVLYINKKIPHRLLKDDLTLDRENLRNHRKLFPPMA
FQASIIAIGAITLQVTKNGLETAIAIAASKTDQLAMFLMNLGLALSTLQNTQNG
AKAYRILEGKDSLIGICIWNSVFAILLIFORRFFSGIFISKADPAVFLAYLYILI
NGSLYWLAILLILRSFIQGLGVFVETLAGIMELVMBAGVAILGLLYFGVGVASAN
PAMWISVLLIPSSLLIKKIKRODNEENLRNL"
BASE COUNT 1158 a 531 c 666 g 1111 t
ORIGIN

Alignment Scores:
Pred. No.: 6,91e-142 Length: 3466
Score: 1695.50 Matches: 312
Percent Similarity: 87.47% Conservative: 51
Best Local Similarity: 75.18% Mismatches: 51
Query Match: 79.04% Indels: 1
DB: 1 Gaps: 1

US-09-469-200D-2 (1-417) x SUB242946 (1-3466)

QY 1 MetArgThrLeuLysAsnLeuLeuThrValValAlaPheSerIlePheTrpValLeuLeu 20
DB 23 ATGGAAAAATATAAAATCTCATTAATTTATGACTTTTATTTCTCGTGGCTCAATTT 82
QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40
DB 83 ATGGGCTTAATGTTTTTGTATTTGGAACTAAGGAAGTCTAACAGTGTATGGGATATT 142
QY 41 LeuIleAlaTyrLeuLeuValLysMetSerLeuSerPhePheTyrLysProPheLysGly 60
DB 143 CTATTAACTATTGTCGATAAAAATGGGATATCTTTTTTTATCGTCCCTATAAAGA 202
QY 61 ArgAlaGlyGlnTyrLysValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSer 80

```

```

|||||
203 AGTGTAGTCAATATAGGTAGCAGCTATTATCCATCTTATATAGAGATGGTGGT 262
QY 81 LeuLeuGluThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVal 100
DB 263 TTAGTAAACTCTAAGAGTGTTCAAAAACAACATATCAATCAATTCGAGAAATTTGCGTA 322
QY 101 ValAspAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAsp 120
DB 323 ATTGACCATGGGTCTAGTAGATAAACAAGGTATATAAATGGTTCGAGAGCTATGTGAAGTTA 382
QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyAsnG 140
DB 383 AATGGC---TTTGAGAGCAACAGTTATCGTTCATCAGATGCTGAAATGTTGGTAAAGA 439
QY 141 HisAlaGlnAlaIlePheGluArgSerAspAlaAspValPheLeuThrValAspSer 160
DB 440 CATGCTCAGGCTTGGCATTTTGAAGGTCTGATGCTGATGTTCTTAAACAGTGGATTC 499
QY 161 AspThrTyrIleTyrProAspAlaLeuGluLeuLeuLysThrPheAsnAspProThr 180
DB 500 GATACCTACATCTATCTCGATGCTCTTGAAGAAATTTATAAGACATTAATGATCCAGAG 559
QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200
DB 560 GTCTACGCTGCAACTGGTCTATTTAATGCAAGAAATAGACAACTAATCTCTTAAGTAGA 619
QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaGlnSerValThr 220
DB 620 CTGACTGATATTCGTTAGTAAATGCAATTTGGTGTAGAACGTGCTGCTCAGTCTGTTACG 679
QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValValPro 240
DB 680 GGAATATTTGGTTGTTCCGGACCTTTAGTATTATAGACGTTCCGTCGGTATTCCTCA 739
QY 241 AsnIleAspArgTyrIleAsnGlnThrLeuGlyIleProValSerIleGlyAspAsp 260
DB 740 AATCTTGAACGGTATATACCTCACAACATTTCTTGGTGTCTGTAAGCATAGGGGATGAC 799
QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLys 280
DB 800 CGTTGTTTGACAAATATGCAACTGATTTGGGAAAAAGCGTTTATCAGTCAACTGCAAGA 859
QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTyrAsn 300
DB 860 TGTGATACTACGTTCCAGATAAGTTTAAAGTTTTCATCAACAACAACAATAATCGTTGGAAT 919
QY 301 LysSerPhePheArgGluSerIleLeuSerValLysLysIleMetAsnAsnProPheVal 320
DB 920 AAGTCATTTTATAGGAGCTTATTATCTCTGTTAAGAGTATTATGACCAACCAAGTGT 979
QY 321 AlaLeuTrpThrIleLeuGluValSerMetPheMetMetLeuValTyrSerValValAsp 340
DB 980 GCTGTTTGGACTATACAGAGTTCCTGATGTCATCATGCTAGTTATTATCTATCTTAGC 1039
QY 341 PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle 360
DB 1040 TTATTGATAGGAGGCTCAAGAATTTAATCTCATAAACTGGTGTGTTTATTGATTATT 1099
QY 361 IlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPhe 380
DB 1100 ATTTCATATAGTACTCTTTGTAGAAATGTTTACATGTTTAAAGCATCCATTTGCTTTT 1159
QY 381 LeuLeuSerProPheTyrGlyValValHisLeuPheValLeuGlnProLeuLysLeuTyr 400
DB 1160 TTATTGTCACCGTTTATGGGATGATACATCTATTCGTTTTCGAACTCTTAAGATAT 1219
QY 401 SerLeuPheThrIleArgAsnAlaAspTrpGlyThrArgLysLys 415
DB 1220 TCGTTTATTACTATAAGAAATGCTACATGGGAACTCGTAAAGAAG 1264
RESULT 6
STRHASAO

```

```

US-09-469-200D-2 (1-417) x STRHASAO (1-1350)

QY      5 LysAsnLeuIleThrValAlaAlaPheSerIlePheTrpValLeuLeuIleTyrValAsn 24
|||    :: ||||:::|||||
Db      91 AAAAAAACITTAATGGTTGTTTATCCTTTTTTTTGATATCATCTGTGATTATCTAAAT 150
|||    :: ||||:::|||||

QY     25 ValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeuLeuIleAlaTyr 44
||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||
Db     151 ATGTATCTATTGGGAACA---TCAACTGTAGGAATTTATGGAGTAATATATAACCCFAT 207
||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||

QY     45 LeuLeuValIlysMetSerLeuSerPhePheTyrLysProPheLysGlyArgAlaGlyGln 64
||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||
Db     208 CTAGTATCAACITGGGATATCTTCCTTTATGAGCCATTTAAGGAAGGCCATGAC 267
||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||

QY     65 TyrLysValAlaAlaIlelleProSerTyrAsnGluAspAlaGluSerLeuLeuGluThr 84
||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||
Db     268 TATAAAGTTGCCTGCTGAATCTCTTTATATCAAGATGCCGAGTCAATTATTAGAACA 327
||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||

QY     85 LeuLysSerValGlnGlnGlnThrTyrProLeuAlaGluIleTyrValValAspAspGly 104
||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||
Db     328 CTTAAAAGTGTGTAGCMAGACCCTATCCGTTATCAGAAATTTATATTGTTGATGATGGG 387
||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||

QY     105 SerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAspThrGlyAspLeu 124
||||   |||||   |||||   |||||   |||||   |||||   |||||   |||||

```

125	QY	SerSerAsnValIleValHisArgSerGluIysAsnGlnGlyIysArgHisIalaGlnIala	144
448	Db	TGTTCGAACGGTTATCGTTACCGTTCCTTGTCAATAAAGGAAACCCGATGCTCAAGCG	507
145	QY	TrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSerAspThrTyrIle	164
508	Db	TGGGCATTGAAGATCTGACGCTGACGTTTTTTTAAACCGTAGATTCAGATACTTATATC	567
165	QY	TyrProAspAlaLeuGluGluLeuLeuLeuIysThrPheAsnAspProThrValPheAlaIala	184
568	Db	TATCCAAATGCCCTTGAAGAAGCTCTTAAAGAGCTTCAATGATGACAGAGTTTATGCTGCA	627
185	QY	ThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIle	204
628	Db	ACAGGACATTTGAATGCTAGNAACACAGAAACTAACTATTAAACGCACTTACAGATATC	687
205	QY	ArgTyrAspAsnAlaPheGlyValGluArgAlaIalaGlnSerValThrGlyIAsnIleLeu	224
688	Db	CGTTACGATAATGCCCTTTGGGTGGAGCGTCTGCTCAATCATTAACAGGTAATATTTTA	747
225	QY	ValCysSerGlyProLeuSerValTyrArgGluValValValProAsnIleAspArg	244
748	Db	GTITGCTCAGGACCATGAGTATTTTATCAGGTGAGTATTTCTTAACTTAGAGCGC	807
245	QY	TyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThr	264
808	Db	TATAAAATCAACATTCCTAGGTTTACCTGTTAGCATTTGGGGATGATCGATGTTTAA	867
265	QY	AsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLysCysIleThrAsp	284
868	Db	AAATTATGCTATTGATTTAGGACCATCTGCTACCAATCAACAGCTAGATGTGACTCAT	927
285	QY	ValProAspLysMetSerTyrTyrLeuLysGlnGlnAsnArgTrpAsnLysSerPhePhe	304
928	Db	GTRACCTTCCAATTAANAAGTTATTAAAGCACAAAACGATGGATAATCTTTTTT	987
305	QY	ArgGluSerIleIleSerValLysLysIleMetAsnAsnProPheValAlaLeuTrpThr	324
988	Db	AGAGAATCTATTATTCTGTGTAAAAAATCTTTCTAAATCCCATCGTTATGAGACT	1047
325	QY	IleLeuGluValSerMetPheMetLeuValTyrSerValValAspPhePheValGly	344
1048	Db	ATTTCGAAGTCGTTATGTTATGATGTGTGATTCGCAATTTGGGAATCTTTTGTTAAT	1107
345	QY	AsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIleIlePheIleVal	364

```

QY 365 AlaLeuCyAsgAnleHisTyrMetLeuLysHisProLeuSerPheLeuLeuSerPro 384
Db 1168 GCTTATGCTGTAATGTCATATATATGCTCAACATCTGCTAGTGTGTTTGTATCTCT 1227
QY 385 PheTyrGlyValLeuHisLeuPheValLeuClpProLeuLysLeuTyrSerLeuPheThr 404
Db 1228 CTGTATGGAATATTACACTGTTGCTCTACAGCCCTAAACTTATTCTTTATGCACC 1287
QY 405 IleArgAsnAlaAspTyrpGlyThrArgLysLysLeu 416
Db 1288 ATTAATAATACGAATGGGACACGTAATAAGGTC 1323

RESULT 7
AE010123 10851 bp DNA linear BCT 03-APR-2002
LOCUS Streptococcus pyogenes strain MGAS8232, section 171 of 173 of the
DEFINITION complete genome.
ACCESSION AE010123 AE009949
VERSION AE010123.1 GI:19749282
SOURCE Streptococcus pyogenes MGAS8232
ORGANISM Streptococcus pyogenes MGAS8232
Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus
1 (bases 1 to 10851)
Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Genome sequence and comparative microarray analysis of serotype M18
Group A Streptococcus strains associated with acute rheumatic fever
outbreaks
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)

JOURNAL
MEDLINE 21927593
PUBMED 11917108
REFERENCE
1 (bases 1 to 10851)
Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.
Direct Submission
Submitted (31-JAN-2002) Laboratory of Human Bacterial
Pathogenesis/Rocky Mountain Laboratories/NIH, 903 S. 4th St.,
Hamilton, MT 59840, USA
FEATURES
source
1. .10851
/organism="Streptococcus pyogenes MGAS8232"
/mol_type="genomic DNA"
/strain="MGAS8232"
/db_xref="taxon:186103"
complement(238..852)
/gene="spyM18_2226"
complement(238..852)
/gene="spyM18_2226"
/gene="spyM18_2226"
/note="best blast match gb|AAK34820.1| (AE006636)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAL98659.1"
/db_xref="GI:19749283"
/translation="MFKKENIKQRYFNFGIVLALITLAIIFAPSSKNADTKSYAKKS
ESKAWTIDKAPKNHAIYKESKEKASIASPIPTVNSVAPTVTEPVPVQEVQ
TVQGVSSVLPNNVLSNGTAGVGSQAQAAATGVPQSTWEHIIARENSNPN
AANASGASGLFTQMPGWSSTATVGDVNAALKAYSAQGLSANGY"
complement(1233..2023)
/gene="spyM18_2228"
complement(1223..2023)
/gene="spyM18_2228"
/note="best blast match gb|AAK34821.1| (AE006636)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"

```

```

/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAL98660.1"
/db_xref="GI:19749284"
/translation="MCKLILGRYIPGDSLHRLDPRSKLLAMIIYIVIFANNVVTN
LMLLFTLAVFLSKIKLSFFLNGVGMIGLILFTLFOMFSGGKVIFSWFISIT
DLGLSQALIFMREVLIIFFSTLLFTPLSLSDAVSLKPLTRKVPABHIGLML
SLSLRFVPTLMDYTRIMNAQRAGVDVFGNLIQKVKSIIPILPLFASFRKRAL
ATAMEARGYQGGRTKYRQLQWLKDSLAIGIVSLGLLFLFKTSL"
complement(2016..2858)
/gene="spyM18_2229"
complement(2016..2858)
/gene="spyM18_2229"
/note="best blast match gb|AAK34822.1| (AE006636)
putative ABC transporter (ATP-binding protein)
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=1
/product="putative ABC transporter (ATP-binding protein)"
/protein_id="AAL98661.1"
/db_xref="GI:19749285"
/translation="MSINLQVSYTYQAGTPFGRALFNINILDLGSYTAFIGHTGS
GKSTLMQLLGLVHPTGTIVSDKODITNHSKKEIKSRKHVGLVQFPESQLEET
VLGDAVGFQNGVSPERAEALAREKIALGVISENFERNFELSGGGMVRAVAIL
AMQPKVLIDEPDTAGLDPKRKELMTIFKLDHSGMTIVLTHLMDVANYADFVIL
DKGLTSGKPKPTIFQOVSLLBPKQLGVPKVTKLAQRLVDRGIPISSLPTILELREY
LRHG"
complement(2834..3724)
/gene="spyM18_2230"
complement(2834..3724)
/gene="spyM18_2230"
/note="best blast match gb|AAK61484.1| (AF082738) ABC
transporter (ATP-binding protein) [Streptococcus
pyogenes]"
/codon_start=1
/transl_table=1
/product="ABC transporter (ATP-binding protein)"
/protein_id="AAL98662.1"
/db_xref="GI:19749286"
/translation="MILKEQAFRLRLNMSATIELKKVTFNYHKDQEKPTLDGVS
FVKQEWLSIIHNGSGSTIRLIDGLLPEPSSGIIYDGLTITNVEIKRITGM
VFQNDPQVATVEDDVAFLGLENKGAHEDIKERNVHALELVGMQNEKEPARLSG
GQKQVATAGVAKMKPILIDENATSMLDPKGLELEIKIKIRDRDQLIVISITHDE
LVALSDRVLVKDGQVESTPEQLFARGDELILQLDLPFTTSVVQMLQEGGYDID
YGLYTELENLQCLLSKM"
complement(3675..4217)
/gene="spyM18_2231"
complement(3675..4217)
/gene="spyM18_2231"
/note="best blast match gb|AAK34824.1| (AE006636)
phosphatidylglycerophosphate synthase [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=1
/product="phosphatidylglycerophosphate synthase"
/protein_id="AAL98663.1"
/db_xref="GI:19749287"
/translation="MIKENIPNLLTVRIAMIPFLFTSSSNKVGWHIFAAVIFAI
ASFTDYLDGLARKHVVSNFGKADPLADKMLVMSAFIMLVGLVLPANVSAVCCR
ELAVTGLRLLVETGGVKVLAAMGLKIKATATQMLSIILLCHWIFLGNVLCLALFFT
IVSGDYKPGASFLFKDFK"
complement(4231..5256)
/gene="spyM18_2232"
complement(4231..5256)
/gene="spyM18_2232"
/note="best blast match gb|AAK61482.1| (AF082738) unknown
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAL98664.1"
/db_xref="GI:19749288"

```



[illegible]

QY 85 LeuLysSerValGlnGlnThrTyrProLeuAlaGluLeuTyrValValAspGly 104  
 Db CTTAAAGTGTGTAGACACACGCTATCCGTTATCAGAAATTTATATTTGATGATGG 429  
 QY 105 SerAlaAspGluThrGlyLeuArgGluAspTyrValArgAspThrGlyAspLeu 124  
 Db AGTTCAACACACATGCAATCAATTAATGAGAGTATGTAATAGACAGATGATAT 489  
 QY 125 SerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArgHisAlaGlnAla 144  
 Db TGTCCAAACGTTATGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTTCCGCTT 490  
 QY 145 TrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSerAspThrTyrIle 164  
 Db TGGGATTTGAAGAATCTGACGCTGAGCTTTTAAACGCTAGACATCTATATC 609  
 QY 165 TyrProAspAlaLeuGluLeuLysThrPheAsnAspProThrValPheAlaAla 184  
 Db TATCCAAATGCTTGAAGAATCTGACGCTGAGCTTTTAAACGCTAGACATCTATATC 610  
 QY 185 ThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIle 204  
 Db ACAGGACATTTGAATGCTTGAAGAATCTGACGCTGAGCTTTTAAACGCTAGACATCTATATC 670  
 QY 205 ArgTyrAspAsnAlaPheGlyValGluArgAlaGlnSerValThrGlyAsnIleLeu 224  
 Db CGTTACGATTAATGCTTGAAGAATCTGACGCTGAGCTTTTAAACGCTAGACATCTATATC 730  
 QY 225 ValCysSerGlyProLeuSerValTyrArgArgGluValValProAsnIleAspArg 244  
 Db GTTTCGTCAGGACATCTGAGTATTTATCGACGCTGAGCTTTTAAACGCTAGACATCTATATC 790  
 QY 245 TyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThr 264  
 Db TATAAATCAACATCTCTAGGTTTACCTGTTAGCTGAGCTTTTAAACGCTAGACATCTATATC 850  
 QY 265 AsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLysCysIleThrAsp 284  
 Db AATTATGCTATTGATTAGGACGCTGCTACCAATCAACAGCTAGATGCTACTGAT 910  
 QY 285 ValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsnLysSerPhePhe 304  
 Db GTACCTTCCAAATTAAGATTTTAAAGCAACAAATCGATGGAATTAATCTTTT 970  
 QY 305 ArgGluSerIleSerValLysLysIleMetAsnAsnProPheValAlaLeuThrTrpThr 324  
 Db ACAGAACTATATTTCTGTTTAAAGAAATCTTCTATATCCATCCATGCTTATGAGCT 1030  
 QY 325 IleLeuGluValSerMetPheMetMetLeuValTyrSerValValAspPheValGly 344  
 Db ATTTTCGAAGTCTGTTATGTTTATGATGTTGATGCGCAATTTGGGAATCTTTTGTAT 1090  
 QY 345 AsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIlePheIleVal 364  
 Db CAAGCTATTCAATAGACCTATTAAACCTTTTTCCTTTTATCCATCATCTTTATCGTT 1150  
 QY 365 AlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPheLeuSerPro 384  
 Db GCTTTATGCTGTAATCTTCAATATATGCTCAACATCTCTGCTAGTTTGTATCTCT 1210  
 QY 385 PheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThr 404  
 Db CRTGATGGAATATTACATGTTGTTGCTTACAGCCCTTAAACCTTTTATCTTTATGAC 1270  
 QY 405 IleArgAsnAlaAspTrpGlyThrArgLysLysLeu 416  
 Db ATTAATAATAGGAATGGGAACACGTAAGAGGTC 1330  
 QY 416 ATTAATAATAGGAATGGGAACACGTAAGAGGTC 1330

RESULT 9

AR232201

LOCUS

DEFINITION

ACCESSION AR232201  
 VERSION AR232201.1 GI:27274189  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 1512)  
 AUTHORS Weigel, P.H., DeAngelis, P.L. and Papaconstantinou, J.  
 TITLE Hyaluronate synthase gene and uses thereof  
 JOURNAL Patent: US 6455304-A 1 24-SEP-2002;  
 FEATURES  
 Location/Qualifiers  
 1. 1512  
 source /organism="unknown"  
 BASE COUNT 484 a 227 c 245 g 556 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.22e-132 Length: 1512  
 Score: 1586.50 Matches: 298  
 Percent Similarity: 85.19% Conservative: 53  
 Best Local Similarity: 72.33% Mismatches: 60  
 Query Match: 73.96% Indels: 1  
 DB: Gaps: 1  
 US-09-469-200D-2 (1-417) x AR232201 (1-1512)  
 QY 5 LysAsnLeuIleThrValValAlaPheSerIlePheTrpValLeuLeuIleTyrValAsn 24  
 Db 205 AAAAAAATCTTAATTTATGCTTTATTTTATGATATCTATCTGATTTATCTAAT 264  
 QY 25 ValTyrPhePheAlaLysGlySerLeuSerIleTyrGlyPheLeuLeuIleAlaTyr 44  
 Db 265 ATGATCTATTGGACAA---TCACCTGATGAGATTTATGGAGTAATATTAATACCTAT 321  
 QY 45 LeuLeuValLysMetSerLeuSerPhePheTyrIleProPheLysGlyArgAlaGlyGln 64  
 Db 322 CTAGTTTCAAACTGGGATATCTTTCTTTTATGAGCAATTTAAAGGAATCCACATGAC 381  
 QY 65 TyrLysValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSerLeuLeuGluThr 84  
 Db 382 TATAAAGTTGCTGTGTAATTTCTTTTATTAAGAGATGCGGATCTATTATTAGAAACA 441  
 QY 85 LeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrValValAspAspGly 104  
 Db 442 CTTAAAGTGTGTAGACACACGCTATCCGTTATCAGAAATTTATTTGTTGATGATGG 501  
 QY 105 SerAlaAspGluThrGlyLysArgIleGluAspTyrValArgAspThrGlyAspLeu 124  
 Db 502 AGTTCAACACACATGCAATCAATTAATTTGAAGATGATGATGATGATGATGAT 561  
 QY 125 SerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArgHisAlaGlnAla 144  
 Db 562 TGTGGAACGTTATCGTTTACCGCTTCCCTTGTCAATAAAGGAACGCCATGCTCAAGCG 621  
 QY 145 TrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSerAspThrTyrIle 164  
 Db 622 TGGGATTTGAAGAATCTGACGCTGAGCTTTTAAACGCTAGACATCTATATC 681  
 QY 165 TyrProAspAlaLeuGluLeuLysThrPheAsnAspProThrValPheAlaAla 184  
 Db 682 TATCCAAATGCTTGAAGAATCTGACGCTGAGCTTTTAAACGCTAGACATCTATATC 741  
 QY 185 ThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIle 204  
 Db 742 ACAGGACATTTGAATGCTTGAAGAATCTGACGCTGAGCTTTTAAACGCTAGACATCTATATC 801  
 QY 205 ArgTyrAspAsnAlaPheGlyValGluArgAlaGlnSerValThrGlyAsnIleLeu 224  
 Db 802 CGTTACGATTAATGCTTGAAGAATCTGACGCTGAGCTTTTAAACGCTAGACATCTATATC 861  
 QY 225 ValCysSerGlyProLeuSerValTyrArgArgGluValValProAsnIleAspArg 244  
 Db 862 GTTTCGTCAGGACATCTGAGTATTTATCGACGCTGAGCTTTTAAACGCTAGACATCTATATC 921

QY	245	TyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThr	264	gene	/db_xref="taxon:160490"
Db	922	TATAAAATCAAACTTCAGTTTACCTGTTAGCATGGGATGCATGTTTAAACA	981	CDS	/note="hasa"
QY	265	AsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLysCysIleThrAsp	284	gene	/note="hasa"
Db	982	AATTAATGCTATGATTAGGACGGCAGCTGCTACCAATCAACAGTACAGTGTGATCTGAT	1041	CDS	/note="Best Blast hit = pir  A53100 hyaluronate synthase A (Hasa) - Streptococcus sp. (group A) >gi 47347 gb AAAI7725.1  (L21187) hyaluronate synthase [Streptococcus pyogenes]"
QY	285	ValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTirpAsnLysSerPhePhe	304	gene	/codon_start=1
Db	1042	GTACCTTCCCAATTAATAAGTATTTAAAGCAACAATCGATGGAATAATCTTTTTT	1101	CDS	/transl_table=11
QY	305	ArgGluSerIleLeuSerValLysIleMetAsnAsnProPheValAlaLeuThrPhe	324	gene	/product="hyaluronate synthase"
Db	1102	AGAGAATCTATTAATTCGTGTAATAAATTCCTTTCTAATCCATCGCTTGCCTTATGACT	1161	CDS	/protein_id="AAK34828.1"
QY	325	IleLeuGluValSerMetPheMetMetLeuValTyrSerValValAspPhePheValGly	344	gene	/db_xref="GI:13623173"
Db	1162	ATTTCGAAGTCGTATGTTATGATGCTGATGTCGCAATCGGAATCTTTTGTTAAT	1221	CDS	/translation="MPFKKTLVLSFLLISLILYLMVLTSTVGVIGVILITVLKGLSLPEKFGPHDYKVAIVPSYNEDAESLFTLKSVAQCTYPLSEIIVDDGSNDTALQLEKGVYVNRVDICRNVIVHRSLYNKKRHAQWAPERSADVFLVSDTYIYNALBELLSKFNFDEFVATGHLNANNKOTNLRLTDIRYDNAPGVVERAOSLTGRLDVGPLSLISYREVLIINPLRYSKQTFGLPVSIGDDRLCTNYAIDLGRTVOSCARCTDVPGLSKYLKQONRWKSKFKESLISVKKLLSNPVALMTIFEVYPMMLI VAIGNLFPNOALQDLIKLFAFLSIIIFVALCRNVHYMIKHPASFLLSPLYGLHLFV LQPLKYSICITKIKTEWGTTRKKVTEK"
QY	345	AsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIlePheIleVal	364	gene	/note="hasa"
Db	1222	CAAGCTATCAATATAGACCTTATTAACACTTTTTCCTTTTATCCATCATCTTATCGTT	1281	CDS	/note="synonym: SPY2201"
QY	365	AlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPheLeuSerPro	384	gene	1541..2749
Db	1282	GCTTATGCTGATGTTATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT	1341	CDS	/gene="hasb"
QY	385	PheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThr	404	gene	/EC_number="1.1.1.22"
Db	1342	CTGTATGGAATATTACACTTGTGCTTACAGCCCTTAAACTTTATCTTTATGACC	1401	CDS	/note="Best Blast hit = sp O07172 UDG_STRPY UDP-GLUCOSE 6-DEHYDROGENASE (UDP-GLC DEHYDROGENASE) (UDP-GLCDH) (UDPGDH) >gi 1075776 pir  A46089 UDPglucose 6-dehydrogenase (EC 1.1.1.22) - Streptococcus pyogenes >gi 8569328 pdb 1DLIA Chain A, The First Structure Of Udp-Glucose Dehydrogenase (Udpqdh) Reveals The Catalytic Residues Necessary For the Two-Told Oxidation >gi 2951891 gb AAA26899.1  (L08444) UDP-glucose hydrogensease [Streptococcus pyogenes]"
QY	405	IleArgAsnAlaAspTrpGlyThrArgGlyLysLeu	416	gene	/codon_start=1
Db	1402	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/product="UDP-glucose 6-dehydrogenase"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/protein_id="AAK34829.1"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/db_xref="GI:13623174"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/translation="MKIAVAGSGVGLSLQNEVTVILPSKVDKINGLSPIQDEIYLLKSKLSIKATLDSKAAYKAEVLATITNTNSRINFTDQHVETVKEVLSYNSHATLIKSTPIGFTETMRQKFTDRIIFSFRESKALDNIPLSRIVTSCENDSPKADAERFALLKSAKNPNVPLINGAEAEVKFALTYALAPRYV NEDTVAESKRNSHMIQIGSYDORIGHNNPFGYGGVCLPKDKLIANTNNIP QTLIEAVSSNNVRKSVIAKQIINVLESPPVGVYRIKMSNSDFESALQDV IDLLKSKDKITIIIVEMLNKLESDQSVLVNDLENFKQANILVTHNRDNLQDVKNK VYSRIQNRQ"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	2931..3845
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/note="hasc"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/note="synonym: SPY2202"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	2931..3845
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/gene="hasc"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/EC_number="2.7.7.9"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/note="Best Blast hit = sp Q54713 HASC_STRPY UDP-GLUCOSE-1-PHOSPHATE URIDYLTRANSFERASE (UDP-GLUCOSE PHOSPHORYLASE) (UDPGP) (ALPHA-D-GLUCOSYL-1-PHOSPHATE URIDYLTRANSFERASE) (URIDINE DIPHOSPHOGLUCOSE PHOSPHORYLASE) >gi 1995767 gb AAA91810.1  (U33452) UDP-glucose pyrophosphorylase [Streptococcus pyogenes]"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/codon_start=1
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/product="UDP-glucose pyrophosphorylase"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/protein_id="AAK34830.1"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/db_xref="GI:13623175"
QY	416	ATTAATAATAGGAATAGGAACACGTAAGGTC	1437	CDS	/translation="MKYRKALIPAGLGTRFLPATYALAKEMLPVDPKPTQIFVEY ALKSGIEELVVTGKAKRSIEDHDSNFEYLNQAKGNELLKVLQDETAINLHFRIT QSHPRGDLAVLOAKAFVGNEDFPVLMGLDMDITNASAKPLTKQIMEDYKTHASTI AVKVPYHEDVSYGVIAPQKAVGLYSVDTEFEKQPQEDAPSDLAIGRTYLLTPRIF GILEQTPGAGNEVOLTAIDTLNKTORVFAFEKNGRYDVGDKFGKMTSIDYALFEV



Db	855	CGTTACGATAATGCTTTGGGTGGAGCGTCTGCTCAATCAATTAACAGGTAATATTTA	914	1.	53896	/organism="Streptococcus pyogenes MGAS315"
QY	225	ValCysSerGlyProLeuSerValThrArgGluValValProAsnIleAspArg	244	/mol_type="genomic DNA"		
Db	915	GIITGCTCAGGACCATGATGATTTATCGAGCGTCAAGTGAATTTCTTAAGTGGAGCGC	974	/strain="MGAS315"		
QY	245	TYVILeAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThr	264	/serotype="M3"		
Db	975	TATAAATCAACAATCTCTAGTTACCTTAGCATGGGATGATCGATGTTTAAACA	1034	/db_xref="taxon:198466"		
QY	265	AsnTyAlaThrAspLeuGlyLysThrValTrpGlnSerThrAlaLysCysIleThrAsp	284	/note="group: A"		
Db	1035	AATTATGCTATTGATTTTAGCAGCGACTGCTTACCAATCAACAGCTAGATGTGATCAT	1094	/note="group: A"		
QY	285	ValProAspLysMetSerThrTrpLeuLysGlnGlnAsnArgTrpAsnLysSerPhePhe	304	/complement(200..1480)		
Db	1095	GTACCTTCCCAATTAAGATGATTTAAGCAACAATCGATGAATGAATATCTTTTTT	1154	/gene="hiss"		
QY	305	ArgGluSerIleLeuSerValLysLysIleMetAsnAsnPropPheValAlaLeuTrpThr	324	/note="synonym: SpyM3_1815"		
Db	1155	AAAGATCTATATTTCTGTTAAATAATCTTTCTAATCCCATCGTTCCTTATGGACT	1214	/protein_id="AAM80422.1"		
QY	325	IleGluGluValSerMetPheMetMetLeuValTrpSerValValAspPhePheValGly	344	/db_xref="GI:21905567"		
Db	1215	ATTTCGAAGTCGTTATGTTTATGATGTTGATTCGCAATTCGGAATCTTTGCTTAAAT	1274	/translation="MKLOPKGTQDILPCDAKQWQVESVARDTFSOVNYGEIPTPMF		
QY	345	AsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIleIlePheIleVal	364	EHEVISRSVGGTDDIVTKEMDFDKGRHILIRPEGAPVVRSYVENKLFAPVQK		
Db	1275	CAGCTATTCAATGACCTTATTAACCTTTTTCCTTTATCCATCATCTTATCGTT	1334	VTUHLGSGSPESAYRQALIDYITPMRDQSKDSORLDENLPIAMAYHFEKLKGLD		
QY	365	AlaLeuCysArgAsnIleHisIleValMetLeuLysHisProLeuSerPheLeuLeuSerPro	384	VERAPSTILDYDEESQAHFAVKAMEALDIPYVIDTNMVRGLDYNHITFEITISVE		
Db	1335	GCCTTATGTCGTAATGTTTATATGATGATCAACATCCTGCTAGTTTTTCTTATCTCT	1394	GSDLTCAGGRVDSLVGYPGPGTGLGRLIMVLEKQITLPIETEMDVILA		
QY	385	PheTyGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTySerLeuPheThr	404	VLGDGANSKALELVAQIRROGTTAERDYLGRKIKAKPKSADTFKAKLVMTLGESEVA		
Db	1395	CTGATGGAATATACACTGTTGTGTCACAGCCCTAAACCTTTATCTTTATGCACC	1454	GKAVINNESQEEVEVSFEDMTNFNANISEQLLS"		
QY	405	IleArgAsnAlaAspTrpGlyThrArgLysLysLeu	416	1700..1882		
Db	1455	ATTAAATACGAATGGGACACGCTAAAGGTC	1490	/gene="rpmf"		
RESULT 11				/note="synonym: SpyM3_1816"		
LOCUS	AE014172	53896 bp	DNA	1700..1882		
DEFINITION	Streptococcus pyogenes MGAS315, section 37 of 37 of the complete genome.			/gene="rpmf"		
ACCESSION	AE014172	AE014074		/note="best blastp hit: gb AAL98635.1  (AE010120) 50S ribosomal protein L32 [Streptococcus pyogenes MGAS8232]"		
VERSION	AE014172.1	GI:21905566		ribosomal protein L32		
KEYWORDS				ribosomal protein L32		
SOURCE	Streptococcus pyogenes MGAS315			/codon_start=1		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.			/transl_table=11		
REFERENCE	1 (bases 1 to 53896)			/product="50S ribosomal protein L32"		
AUTHORS	Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.			/protein_id="AAM80424.1"		
TITLE	Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence			/db_xref="GI:21905569"		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)			/translation="MRVNITLHKESGERLYLTSKNKNRPDRILQKKYSPKLRKHYT FTEVK"		
REFERENCE	2 (bases 1 to 53896)			1898..2047		
AUTHORS	Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.			/note="synonym: SpyM3_1817"		
TITLE	Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence			1898..2047		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)			/gene="rpmf"		
REFERENCE	2 (bases 1 to 53896)			/note="best blastp hit: gb AAK34796.1  (AE006634) 50S ribosomal protein L33 [Streptococcus pyogenes M1 GAS], and gb AAL98636.1  (AE010120) 50S ribosomal protein L33 [Streptococcus pyogenes MGAS8232]"		
AUTHORS	Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.			/codon_start=1		
TITLE	Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence			/transl_table=11		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)			/product="50S ribosomal protein L33"		
REFERENCE	2 (bases 1 to 53896)			/protein_id="AAM80425.1"		
AUTHORS	Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.			/db_xref="GI:21905570"		
TITLE	Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence			/translation="MRVNITLHKESGERLYLTSKNKNRPDRILQKKYSPKLRKHYT FTEVK"		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)			1898..2047		
REFERENCE	2 (bases 1 to 53896)			/note="synonym: SpyM3_1818"		
AUTHORS	Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.			2328..2942		
TITLE	Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence			/gene="cadd"		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)			/note="synonym: SpyM3_1815"		
REFERENCE	2 (bases 1 to 53896)			/note="best blastp hit: gb AAK34797.1  (AE006634) putative cadmium resistance protein [Streptococcus pyogenes M1 GAS]"		
AUTHORS	Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.			/codon_start=1		
TITLE	Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence			/transl_table=11		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)			/product="putative cadmium resistance protein"		
REFERENCE	2 (bases 1 to 53896)			/protein_id="AAM80425.1"		
AUTHORS	Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.			/db_xref="GI:21905570"		
TITLE	Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence			/translation="MIONVVTSLYSGTAVDLLLIIMLFFFAKRSKREDIINLYLQGF LGSVLLISLLFAFVLDYTPSKELIGLLGLLPIFGLKVLGLGSDGEAIKAEGLSK DNKNLIFLIVAMITFASCGADNIGVFPVFTTILNLANLIVALLFLVMIYLLVLSAOKL		

```

gene
2954. .3292
/sequence="cadx"
/notes="synonym: SpyM3_1819"
CDS
2954. .3292
/sequence="cadx"
/notes="best blastp hit: gb|AAK34798.1| (AE006634) putative
cadmium efflux system accessory protein [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative cadmium efflux system accessory
protein"
/protein_id="AAK0426.1"
/db_xref="GI:21905571"
/translation="MKKDSICQGVINQONVYATNLYEKEKVKQSLKILSKFTDNKQ
INIFILAVEELCVDCIACLLNLSMASHLRLKLANILDTREGKIIYFIKDE
EIRDFNOLG"
3343. .4266
/sequence="SpyM3_1820"
3343. .4266
/sequence="SpyM3_1820"
/notes="best blastp hit: gb|AAK34799.1| (AE006634)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK0427.1"
/db_xref="GI:21905572"
/translation="MKFEEYSGEIPTRIKNLRKKNITQSELNAGQVQVESKRR
PITSVNLALNATSSYIVFEGDEFIENLFYFFSILYROLEAVDENLYSFM
DLSIQSSCLSIATKAFNFIQKRFMISTEMDFHKDDIDVWGGESYNPARS
FRNPINELTVDFDEMAIDILLMLRNLIRSFNVCNLTFLDKNGEPTFLDKI
DSINIKWSENVSTELPNLIKLRNLPNIGPMWVNDILERMKYEDIPKSLYLSVPL
VSKKARSLSLSKNSGSKDKSKFAQLSL"
4332. .5084
/sequence="SpyM3_1821"
4332. .5084
/sequence="SpyM3_1821"
/notes="best non-GAS blastp hit: gb|AAK79835.1|AE007695_8
(AE007695) FltK-like DNA segregation ATPase, YDCQ
B. subtilis ortholog [Clostridium acetobutylicum]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK0428.1"
/db_xref="GI:21905573"
/translation="MAIPDLKESNLQOQKIPVIKHFLNLDKSHWAIAGNSGSK
PYALTFLSVLPMSEGLIIDPKFPPSQWAKENKIAVHIPVENHSKSDFYVQVNEQL
NQCATLQKRAIYDNPNOHFTLTVIDEVLSALSEGYNKNTKEAFFSLISLALIG
HATKILFLGSGQRFQDHTIPISVREQLNVLQIGNINOKTQFLPDLDPBGIVPTG
HGTGILQVVDNHSYQVLPILLCTPYTKRGIL"
5081. .5683
/sequence="SpyM3_1822"
5081. .5683
/sequence="SpyM3_1822"
/notes="best blastp hit: gb|AAK34801.1| (AE006634)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK0429.1"
/db_xref="GI:21905574"
/translation="MTTKVFMRSFNLIYLLIAPICVYVFGMLFNLYSVOLSTVTFV
NLDTKVNTDTLSSVCGFALISILLALISGLMKIQLRNTDLSGNVCOSIYRTLLRH
FLFQREKVNSEHFQVTTINPIINGENRNVKCIIVDIRKDAITVFLKVPQQGQ
KILKDLVQLKEEVIQHTDIYFSSPVRVNMJFIGQR"
6159. .6323
/sequence="SpyM3_1823"
6159. .6323
/sequence="SpyM3_1823"
/notes="best blastp hit: gb|AAK34556.1| (AE006610)
hypothetical protein [Streptococcus pyogenes M1 GAS]"

```

```

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK0430.1"
/db_xref="GI:21905575"
/translation="MLFGTLLIAISYIPMSLWGWGYPVAILWLIIMRSALRIERF
KKTVDVQTY"
complement(6651..6809)
/sequence="SpyM3_1824"
complement(6651..6809)
/sequence="SpyM3_1824"
/notes="best blastp hit: emb|CAK65345.1| (Y18538) PR7
protein [Plasmodium falciparum]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK0431.1"
/db_xref="GI:21905576"
/translation="MFLINNTNFEDIFLKHVFVKSSLIVKVNKIYCKCKPYDNY
QKDKRKA"
6894. .7736
/sequence="SpyM3_1825"
6894. .7736
/sequence="SpyM3_1825"
/notes="best non-GAS blastp hit: gb|AAK99693.1| (AE008462)

Alignment Scores:
Pred. No.:      2,59e-130      Length:      53896
Score:          1585.50      Matches:      298
Percent Similarity: 85.19%      Conservative: 53
Best Local Similarity: 72.33%      Mismatches: 60
Query Match:      73.92%      Indels: 1
DB:              1          Gaps: 1

US-09-469-200D-2 (1-417) x AE014172 (1-53896)

QY 5 LysAsnLeuIleThrValValAlaPheSerIlePheTrpValLeuLeuIleTyrValAsn 24
Db 33885 AAAAACAATTAAATGGTTTATTCCTTTATGTTGATATCTATCTGATTATCAAT 33944
QY 25 ValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeuLeuIleAlaTyr 44
Db 33945 ATGTATCTATTGGACAA---TCAACTGTAGGAATTTATGGAGTAAATTAATCACTAT 34001
QY 45 LeuLeuValLysMetSerLeuSerPhePheTyrLysProPheLysGlyValGlaGlyGln 64
Db 34002 CTAGTATTAACTGGATATCTTTCCTTATGAGCAATTAAGGAAATCCACATGAC 34061
QY 65 TyrLysValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSerLeuLeuIleThr 84
Db 34062 TATAAAGTTGCTGCTGTAATTCCTTCTTAAATCAAGATCCCGAGTCATATTAGAAACA 34121
QY 85 LeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrValValAspAspGly 104
Db 34122 CTTAAAGTGTGTTAGCACAGACCTATCCGTTATCAGAAATTTATATTGTTGATGATGG 34181
QY 105 SerAlaAspGluThrGlyIleLysArgGileGluAspTyrValArgAspThrGlyAspLeu 124
Db 34182 AGTTCACACACAGATGCAATACAAATTAATTAAGAGTATGTAATAGAGAGTGGATATT 34241
QY 125 SerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArgHisAlaGlnAla 144
Db 34242 TGTCAAGACGTTATGTTTCACCGTTCCCTTGTCAATAAAGAAACGCCATGCTCAAGG 34301
QY 145 TrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSerAspThrTyrIle 164
Db 34302 TGGGCAATTTGAAGATCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 34361
QY 165 TyrProAspAlaLeuGluLeuLeuLysThrPheAsnAspProThrValPheAlaAla 184
Db 34362 TATCCAAATGCCCTAGAGAGACTCTCTTAAAGAGTCAATGATGATGATGATGATGATG 34421
QY 185 ThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIle 204

```

```
Db 34422 ACAGGACATTTGATGCTAGAAACAGACAACAACTAACTATTATPACGGGACCTTACAGATATC 34481
QY 205 ArgTyrAspAsnAlaPheGlyValGluArgAlaGlnSerValThrGlyAsnIleLeu 224
Db 34482 CGTTAGCATAATGCTTTGGGTGGAGCGTCTGCTCAATCATTAACAGGATATATTTA 34541
QY 225 ValCysSerGlyProLeuSerValTyrArgGluValValProAsnIleAspArg 244
Db 34542 GTTGTCTCAGGACCAATGAGTATTATCGAGTGAAGTGAATTAATTCCTTACCTTAGAGCGC 34601
QY 245 TyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspArgCysLeuThr 264
Db 34602 TATAAAATCAACATTCCTAGTTCCTTACCTTAGGATGGGATGATCGATGTTTAAACA 34661
QY 265 AsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLysCysIleThrAsp 284
Db 34662 AATATGCTATTATTTAGAGCGCTCTCTACCAATCAACAGCTAGATGTGATACAT 34721
QY 285 ValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTyrAsnLysSerPhePhe 304
Db 34722 GTACCTTCCCAATTAAGATTAATTAAGCAACAATAATCGATGGAATTAATCTTTT 34781
QY 305 ArgGluSerIleLeuSerValLysLysIleMetAsnAsnProPheValAlaLeuThrThr 324
Db 34782 AGAATCTATTATTTCTGCAAAAAAATCTTCTTAATCCCATCGTTGCTTATGGACT 34841
QY 325 IleGluGluValSerMetPheMetMetLeuValTyrSerValValAspPheValGly 344
Db 34842 ATTTTCGAAGTCGTTATGTTATGATGTTGATGTCGAATGGGAATCTTTTGTATA 34901
QY 345 AsnValArgGluPheAspTyrLeuArgValLeuAlaPheLeuValIlePheIleVal 364
Db 34902 CAAGCTATTCAATAGACCTTATAAAGCTTTTGCCTTTTATCCATCACTTATCGTT 34961
QY 365 AlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPheLeuLeuSerPro 384
Db 34962 GCTTTATGTCGTAATGTTTATGATGTTTATGATGTTTATGATGTTTATGATGTT 35021
QY 385 PheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThr 404
Db 35022 CTGTATGGAATTAATACCTTGTGTGCTTACAGCCCTTAAACTTATTTATGATGACC 35081
QY 405 IleArgAsnAlaAspTyrGlyThrArgLysLysLeu 416
Db 35082 ATTAAAAACCGAATGGGACACGCTAAAGGTC 35117

RESULT 12
AP005146 323825 bp DNA linear BCT 13-JUN-2003
LOCUS Streptococcus pyogenes SSI-1 DNA, complete genome, section 6/6.
ACCESSION AP005146 BA000034
VERSION AP005146.1 GI:28811737
KEYWORDS Streptococcus pyogenes SSI-1
SOURCE Streptococcus pyogenes SSI-1
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Nakagawa, I., Kurokawa, K., Yamashita, A., Nakata, M., Tomiyasu, Y.,
Okahashi, N., Kawabata, S., Yamazaki, K., Shiba, T., Yasunaga, T.,
Hayashi, H., Hattori, M., and Hamada, S.
TITLE Genome Sequence of an M3 Strain of Streptococcus pyogenes Reveals a
Large-Scale Genomic Rearrangement in Invasive Strains and New
Insights into Phage Evolution
JOURNAL Genome Res. 13 (6), 1042-1055 (2003)
MEDLINE 22683278
REFERENCE 2 (bases 1 to 323825)
AUTHORS Yamashita, A., Nakagawa, I., Kurokawa, K., Nakata, M., Tomiyasu, Y.,
Yamazaki, K., Okahashi, N., Kawabata, S., Yasunaga, T., Hattori, M.,
Hayashi, H., and Hamada, S.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) Ken Kurokawa, Osaka University, Genome
```

```
Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail:ken@gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
Fax:81-6-6879-2047)
This clone was isolated from a patient presenting with toxic shock
like syndrome.
COMMENT genome project
FEATURES
source Location/Qualifiers
1. 323825
/organism="Streptococcus pyogenes SSI-1"
/mol_type="genomic DNA"
/strain="SSI-1"
/db_xref="taxon:193567"
complement(422..1012)
/gene="SPS1572"
complement(422..1012)
/gene="SPS1572"
/notes="similar to GB:AAK33432.1 (AE006502) percent
identity 100 in 196 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative ATP-dependent protease proteolytic
subunit"
/protein_id="BAC64667.1"
/db_xref="GI:28811738"
/translation="MIPVIEQTSRGERSYDIYSRLKKDRILMTGPVEDNMANSVIA
QLFLDAQNTKDIYLVNTPGGSVAGLAIVDMFIRADYQTIIVMGMAASGTVIA
SSGTGKRPMLPNAEYMIHQPMTGGTQOTDMAAEHLKTHRLKLEILAQNAKG
TIQIHKAERDYWMAEETLAYGFIIDIMENNELK"
complement(1235..1864)
/gene="SPS1573"
complement(1235..1864)
/gene="SPS1573"
/notes="similar to GB:AAK33430.1 (AE006502) percent
identity 100 in 209 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative uracil phosphoribosyltransferase"
/protein_id="BAC64668.1"
/db_xref="GI:28811739"
/translation="MGKQVISHPLQHKLSILRQRTTSTKDFRELVNEIAMLMGYEV
SRDLPLEVDIQTVPSTVOKLAGKLAIVPLIRAGIGMVDGLLSIVPAKVGHTGM
YRNEVLEPVEVLKLPEDINORQIFLVDPLMATGGSATLAVDSLKKEGANIKFVCL
VAAEGVKLQEAHPDIDIFTAALDHLNHEGYIVPGLGDAGDRFGTK"
complement(2032..3666)
/gene="SPS1574"
complement(2032..3666)
/gene="SPS1574"
/notes="similar to GB:AAK33429.1 (AE006501) percent
identity 99 in 544 aa"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC64669.1"
/db_xref="GI:28811740"
/translation="MSTKKQLTQELVQGAHSTAGNISRLGLVLIIPVYIMWG
QYANALFNNGINVAIFLLSTTGLNVAIAQVAKYNSMGQTEHSYQLIRSLTKL
MLGLIFSAIMYLGSPFASLGGDDTLVPMHSLSLAVFFPVNSVIRGIFQGHNN
IKPYASQIAEQLIRVIMLLTITLMKLGSDYASAVTQSTFAAFIGWASMGVLGY
YLMKQILAAIFSKPDPHTVIDIKGLLLETKEISIPITVTSIAQIFOLDIMTFTVT
MTLFTDYSQSLVILFGYFNANPAKTWVLIAVAASIGVGIALLTENVYKDKMAA
RLINNIEMLYNFPALTCAIILARPLYSVFYGCASEDAIRHLFVAVLFTQLMALYT
LFSPMLQALFENRRAIYFAYGILKVLQIPLIYLHAYGLPALTIALVVPVILMY
RLYQVTHENRKLQLLTETLLMGLVVFVANNLLGYAFKPTGLTSLTLLII
GGGMVTVYATLTLLTHQLDKLGSKASRLQKLGW"
3754..5199
/gene="SPS1575"
3754..5199
/gene="SPS1575"
/notes="similar to GB:AAK33428.1 (AE006501) percent
```

Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)

COMMENT genome project

This clone was isolated from a patient presenting with toxic shock like syndrome.

FEATURES Location/Qualifiers

source

1. 323825

/organism="Streptococcus pyogenes SSI-1"

/mol\_type="genomic DNA"

/strain="SSI-1"

/db\_xref="taxon:193567"

complement(422..1012)

/gene="SPS1572"

complement(422..1012)

/gene="SPS1572"

/notes="similar to GB:AAK33432.1 (AE006502) percent

identity 100 in 196 aa"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="putative ATP-dependent protease proteolytic

subunit"

/protein\_id="BAC64667.1"

/db\_xref="GI:28811738"

/translation="MIPVIEQTSRGERSYDIYSRLKKDRILMTGPVEDNMANSVIA

QLFLDAQNTKDIYLVNTPGGSVAGLAIVDMFIRADYQTIIVMGMAASGTVIA

SSGTGKRPMLPNAEYMIHQPMTGGTQOTDMAAEHLKTHRLKLEILAQNAKG

TIQIHKAERDYWMAEETLAYGFIIDIMENNELK"

complement(1235..1864)

/gene="SPS1573"

complement(1235..1864)

/gene="SPS1573"

/notes="similar to GB:AAK33430.1 (AE006502) percent

identity 100 in 209 aa"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="putative uracil phosphoribosyltransferase"

/protein\_id="BAC64668.1"

/db\_xref="GI:28811739"

/translation="MGKQVISHPLQHKLSILRQRTTSTKDFRELVNEIAMLMGYEV

SRDLPLEVDIQTVPSTVOKLAGKLAIVPLIRAGIGMVDGLLSIVPAKVGHTGM

YRNEVLEPVEVLKLPEDINORQIFLVDPLMATGGSATLAVDSLKKEGANIKFVCL

VAAEGVKLQEAHPDIDIFTAALDHLNHEGYIVPGLGDAGDRFGTK"

complement(2032..3666)

/gene="SPS1574"

complement(2032..3666)

/gene="SPS1574"

/notes="similar to GB:AAK33429.1 (AE006501) percent

identity 99 in 544 aa"

/codon\_start=1

/evidence=not\_experimental

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="BAC64669.1"

/db\_xref="GI:28811740"

/translation="MSTKKQLTQELVQGAHSTAGNISRLGLVLIIPVYIMWG

QYANALFNNGINVAIFLLSTTGLNVAIAQVAKYNSMGQTEHSYQLIRSLTKL

MLGLIFSAIMYLGSPFASLGGDDTLVPMHSLSLAVFFPVNSVIRGIFQGHNN

IKPYASQIAEQLIRVIMLLTITLMKLGSDYASAVTQSTFAAFIGWASMGVLGY

YLMKQILAAIFSKPDPHTVIDIKGLLLETKEISIPITVTSIAQIFOLDIMTFTVT

MTLFTDYSQSLVILFGYFNANPAKTWVLIAVAASIGVGIALLTENVYKDKMAA

RLINNIEMLYNFPALTCAIILARPLYSVFYGCASEDAIRHLFVAVLFTQLMALYT

LFSPMLQALFENRRAIYFAYGILKVLQIPLIYLHAYGLPALTIALVVPVILMY

RLYQVTHENRKLQLLTETLLMGLVVFVANNLLGYAFKPTGLTSLTLLII

GGGMVTVYATLTLLTHQLDKLGSKASRLQKLGW"

3754..5199

/gene="SPS1575"

3754..5199

/gene="SPS1575"

/notes="similar to GB:AAK33428.1 (AE006501) percent

[illegible]

```

QY 165 TyrProAspAlaLeuGluLeuLeuLysThrPheAsnAspProThrValPheAlaAla 184
DB 304322 TATCAATGCCCTTGAAGAACTCTAAAGCTTCAATGATGACAGCTTATCTGCGCA 304381
QY 185 ThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIle 204
DB 304382 ACAGGACATTGGATGCTAGAACAGACAACTATCTATTAAAGCGCTTACAGATATC 304441
QY 205 ArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThrGlyAsnIleLeu 224
DB 304442 CGTTACGATAATGCCCTTTGGGTGGAGCTGCTGCTCAATCATTAACAGGTAATATTTA 304501
QY 225 ValCysSerGlyProLeuSerValTyrArgArgGluValValProAsnIleAspArg 244
DB 304502 GTTGTCTCAGGACCATGAGTATTTATCGACGTGAGTGATATTCCTCACTTACAGGCGC 304561
QY 245 TyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThr 264
DB 304562 TATATAAATCAACATCTCTAGCTTACCTGTTAGCATGGGATGATGATGTTTAAACA 304621
QY 265 AsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLysCysIleThrAsp 284
DB 304622 AATTATGCTATTGATTAGACGCACTGCTACCAATCAACAGCTAGATGATGATGATGAT 304681
QY 285 ValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsnLysSerPhePhe 304
DB 304682 GTACCTTTCCAAATAAAGTTATTAAACCAACAAATCGATGGAATAAATCTTTT 304741
QY 305 ArgGluSerIleLeuSerValLysLysIleMetAsnAsnPropheValAlaLeuThrPhe 324
DB 304742 AGAGAATCTATTATCTCTCAAAAAATCTTCTTAATCCCATGCTGCTTATGAGCT 304801
QY 325 IleLeuGluValSerMetPheMetLeuValTyrSerValValAspPheValGly 344
DB 304802 ATTTCCGAAGTCGTATGTTATGATGTTGATGTCGCAATGCGGAATCTTTTGTGTTAAT 304861
QY 345 AsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIleLeuPheIleVal 364
DB 304862 CAAGCTATTCAATGACCTTATTAACTTTTGGCTTTTATCCATCATCTTATCGTT 304921
QY 365 AlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPheLeuLeuSerPro 384
DB 304922 GCTTTATGCTGTAATGTTCTATATATGTCACCAATCTGCTGCTGTTTATGTTATCCT 304981
QY 385 PheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThr 404
DB 304982 CTGTATGGAATATTACACTGTTGTTCTTACAGCCCTTAAACCTTATCTTTATGTCACC 305041
QY 405 IleArgAsnAlaAspTrpGlyThrArgLysLysLeu 416
DB 305042 ATTAAAAATACGAATGGGACACGTAAGATC 305077

RESULT 13
AF065404
LOCUS
DEFINITION Bacillus anthracis virulence plasmid pX01, complete sequence.
ACCESSION AF065404
VERSION AF065404.1 GI:4894216
KEYWORDS
SOURCE Bacillus anthracis
ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
cereus group.
REFERENCE
1 (bases 1 to 181654)
AUTHORS Okinaka,R.T., Cloud,K., Hampton,O., Hoffmaster,A.R., Hill,K.K.,
Keim,P., Koehler,T.M., Lamke,G., Kumano,S., Mahillon,J., Manter,D.,
Martinez,Y., Rieke,D., Svensson,R. and Jackson,P.J.
TITLE Sequence and organization of pX01, the large Bacillus anthracis
plasmid harboring the anthrax toxin genes
JOURNAL J. Bacteriol. 181 (20), 6509-6515 (1999)
MEDLINE 99445483
PUBMED 10515943
REFERENCE
2 (bases 1 to 181654)

```

Okinaka,R.T., Cloud,K., Hampton,O., Hoffmaster,A., Hill,K., Keim,P.,  
 Koehler,T., Kumano,S., Lamke,G., Manter,D., Martinez,Y.,  
 Rieke,D.O., Svensson,R. and Jackson,P.J.  
 Direct Submission  
 Submitted (14-MAY-1998) Life Sciences Division, Los Alamos National  
 Laboratory, P43, LS-6, HRL-1, MS M888, Los Alamos, NM 87545, USA  
 Location/Qualifiers  
 source  
 1..181654  
 /organism="Bacillus anthracis"  
 /mol\_type="other DNA"  
 /strain="Sterne"  
 /db\_xref="taxon:1392"  
 /plasmid="virulence plasmid pX01"  
 complement(1127..1630)  
 /note="Hypothetical protein, Methanobacterium  
 thermoautotrophicum, (AB000831)"  
 /codon\_start=1  
 /transl\_table=11  
 /product="pX01-01"  
 /protein\_id="AAD32305.1"  
 /db\_xref="GI:4894217"  
 /translation="MEVLIPELILIAVLIPNSVKKHVPKWKAGKELIYRMLSKL  
 DKPSYLVHNVYTVETGTDIDHIVIAETGVVYVTKNYEGWIGSEKAAARWQGI  
 FRKSFQNFPHONYKHIAEWLIEQQLFCISMAFPHKGLARVNVHSEKHKYKL  
 HNKFAQ"  
 complement(1639..1644)  
 RBS 2751..2755  
 RBS 2760..3338  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="pX01-02"  
 /protein\_id="AAD32306.1"  
 /db\_xref="GI:4894218"  
 /translation="MLEHTLPSLVQLVPLIVSTGIYILLELDTENKWTESYGF  
 AILITLRFQFNKSLISLQHEKFWKSLNIFLVSILVLFASIFTLKLAFLQW  
 TWILFELCLIDRLPHNLYRVKEIICDKKAKIERIIPNNEEHGTHITKTRDEQ  
 YFLYAVYLSGLISPSIEEFVEYERNLMGR"  
 complement(3389..4024)  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="pX01-03"  
 /protein\_id="AAD32307.1"  
 /db\_xref="GI:4894219"  
 /translation="MKREKMGIVLCTILVILLVPLQAQASVYDFNKLKGDVNS  
 SPKGLANRKTTPKDDSGSSIIOKEGLYAPAFKGVNFISSIFVLGIIVMVAL  
 IYKNGQWKATGTFSPISLMSNRLSVLPSDFSNFIITDFLDFKATALIF  
 TFMLLAGLRRLRHSTKQPIYRWANRITGSCILCLMLAVLAPWLFANN"  
 complement(4024..4653)  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="pX01-04"  
 /protein\_id="AAD32308.1"  
 /db\_xref="GI:4894220"  
 /translation="MNINSNNVKELEQTVERKKQVQLRVSKSSNIHAYMIFP  
 EFSEVKNTLEDYIENLSTVKYIDRLKSQOEIEKQNHSEFIMTCEVEDGVIDDE  
 THFTVTKPTTANKVFLKGLNGLSLKELHITVNNNINRIEIFYENETTY  
 VAYKEITANIMGAYTKANIYAADLYAAEGNEIKLKSIPKLAGDV"  
 complement(4034..4039)  
 RBS complement(4660..4665)  
 RBS complement(4989..5309)  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="pX01-05"  
 /protein\_id="AAD32309.1"  
 /db\_xref="GI:4894221"  
 /translation="WELLEYDMYQQYVSIPPETPVFPQQRVADVMDQDFVWRT  
 TKYVEQGLTRVNMNSNKRWIHTGEILPENQLLYKLEVDSPSYITLVESNRO  
 LGNI"  
 complement(5318..5323)  
 RBS complement(5720..6220)  
 CDS  
 /codon\_start=1  
 /transl\_table=11  
 /product="pX01-06"



```
QY 127 -----AsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140
Db 111593 GAAATATTAGTATTCCTGACTTAATCGTACATGTTTACCTAAGAAATGCGGGAAAGA 111652
QY 141 HisAlaGlnAlaTrrAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160
Db 111653 CATGCTCAATATTGGCGTTTAAACGACACACAGATGCTATTGTTACCATGATTCA 111712
QY 161 AspThrTyrIleTyrProAspAlaLeuGluLeuLeuLysThrPheAsnAspProThr 180
Db 111713 GATGCTGATTGTTCCCAATGCTGTAGAGATTATTGAAACCCCTTATGATGATGAAAA 111772
QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnThrArg 200
Db 111773 GAAATGCGCAACTGCTGACGTGCAATTCGTAATGAGAAATGATAATTATTATACAAA 111832
QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThr 220
Db 111833 CTAATTGATATGCGTTATGACAAATGCGTTCGTTGGAGCGTGCAGACACAGTCGTAACA 111892
QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgArgGluValValPro 240
Db 111893 GGAATATGTTCTGTTGTTAGTGGCGCTTAAGTTGTTATCGTAGAAGTAATAACTGAA 111952
QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260
Db 111953 AATTAGAACATTATGGAAGTCAGATGTTCTTGGTGGAGGAGTGCAGTTTGGACATGAT 112012
QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLys 280
Db 112013 AGATGCTAACTAATTATGCTATTTTGAAAGGAAACAGTTTATCAATCCACTGCTCGA 112072
QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsn 300
Db 112073 TGTATTACTGATGCTCAACTACATTAAACAATTTCTTAAACAGCACTACGTTGGAAC 112132
QY 301 LysSerPhePheArgGluSerIleSerValLysLysIleMetAsnAsnProPheVal 320
Db 112133 AGTCATTTTTAGAGAAAGTTAATTTACCTGGCATGGGTATGATGAAAAACCAATGTT 112192
QY 321 AlaLeuThrThrIleLeuGluValSerMetPheMetLeuValTyrSerValValAsp 340
Db 112193 CTGTGTTGCAATTTTCGAAATATCGTTATGAGATTTATTTGGGCTTCCCTACTCTTA 112252
QY 341 PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle 360
Db 112253 AGPATTATTCTCAAGCAAGTCATGATAGGTTAATTTTGGCTGTTATTATTGCGTTAT 112312
QY 361 IlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPhe 380
Db 112313 ATTTCAATAGCTGATATGCTAGAAATGATTTATCTATTAACATCCCTTACTTTC 112372
QY 381 LeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyr 400
Db 112373 TTACTGGCGCAATTATGGAATCTCCATGATTTACCATATTACCTATACGCTTTTAT 112432
QY 401 SerLeuPheThrIleArgAsnAlaAspTrpGlyThrArg 413
Db 112433 GCTTTACTAATCTAATAATCTAATGTTGGGGAACAGT 112471

RESULT 14
LOCUS AE011190 181677 bp DNA circular BCT 14-JUN-2002
DEFINITION Bacillus anthracis str. A2012 plasmid pX01, complete sequence.
ACCESSION AE011190
VERSION AE011190.1 GI:20520075
KEYWORDS
SOURCE
ORGANISM Bacillus anthracis str. A2012
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
cereus group.
REFERENCE 1 (bases 1 to 181677)

AUTHORS Read,T.D., Salzberg,S.L., Pop,M., Shumway,M., Umayam,L., Jiang,L.,
Holtzapple,E., Busch,J.D., Smith,K.L., Schupp,J.M., Solomon,D.,
Keim,P. and Fraser,C.M.
TITLE Comparative genome sequencing for discovery of novel polymorphisms
in Bacillus anthracis
JOURNAL Science 296 (5575), 2028-2033 (2002)
MEDLINE 22061436
PUBMED 12004073
REFERENCE 2 (bases 1 to 181677)
AUTHORS Read,T.D., Salzberg,S.L., Pop,M., Shumway,M., Umayam,L., Jiang,L.,
Holtzapple,E., Busch,J.D., Smith,K.L., Schupp,J.M., Solomon,D.,
Keim,P. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..181677
/organism="Bacillus anthracis str. A2012"
/mol_type="genomic DNA"
/strain="A2012"
/db_xref="taxon:191218"
/plasmid="pX01"
/notes="Florida strain"
462..554
/gene="BXA0001"
462..554
/gene="BXA0001"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA025958.1"
/db_xref="GI:20520076"
/translation="MVARILHLIKFKVVKSKRHSNECLFLWTK"
complement(96)..1109
/gene="BXA0002"
complement(96)..1109
/gene="BXA0002"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA025959.1"
/db_xref="GI:20520077"
/translation="MLYVLCFLVISIVAILNKKYKLGKNGKDFVAKNG"
complement(1127)..1765
/gene="BXA0003"
complement(1127)..1765
/gene="BXA0003"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX01-01)"
/protein_id="AA025960.1"
/db_xref="GI:20520078"
/translation="MEYLFELILAVILPNSVKKHVPKWKAGKEKLVKRLSKL
DPKSYVLENNVTYETGDTQIDHIVIAETGVVETKNGWYIGSEKAAHWTOGI
FRKSFQNPQPHQVKKHIALEWLEQOLPCISMAAFKPKSLKRVNVHSEKHHVLY
NDLQKLESYTDVOLTNDVEVQHIYHTILRANIMDKIEKKHKVYLHNKFAQ"
complement(1782)..2165
/gene="BXA0004"
complement(1782)..2165
/gene="BXA0004"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA025961.1"
/db_xref="GI:20520079"
/translation="MNRNREYIALRNISISIGLWLTCTICLSIYFSSVYLLIGCFIL
SLKYGIMGVKPKVNEKRLPTWANGVLEKPNRRVTRFPQFRGLLIVSLALGFEMVYK
GLCISFIIGLGETHVFLMRNRNM"
2291..2875
gene
```

```

/gene="BXA0005"
2291. .2875
/gene="BXA0005"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM25962.1"
/db_xref="GI:20520080"
/translation="MLKLVVNTGTEPEGSFSAFNTFTKDCVYLSKSDSLSQALAA
DSNNPLVYECGNEETYEIGTEDPLNSHIIHQOLREKQAIIAESAIPAQSIDPLHKD
AIWGSVNYGWIINDSKKRFALAQVFPFGDNNSSKKVASFPFLRDHGAKGIRHYM
CWVNEKGIGVEHLLIGGVIMKTAAGVKKLGFN"
2895. .3473
/gene="BXA0006"
2895. .3473
/gene="BXA0006"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX01-02)"
/protein_id="AAM25963.1"
/db_xref="GI:20520081"
/translation="MLSHLTIISLVQLVPLVIVSGHIIHLELTDTEKWTFSYEGIF
AIFLTRQVQNKKSIIISLQNHKEVSKLPIVSLFELFASITPLTKAFILQM
TVILFELYLCYIDRLPIHNLVYKEIICDKQKLIERIIPNEEHGTHITRDEQ
YFLYTVLSGLSPISSEVEFEYERNLMGR"
complement(3524. .4111)
/gene="BXA0007"
complement(3524. .4111)
/gene="BXA0007"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX01-03)"
/protein_id="AAM25964.1"
/db_xref="GI:20520082"
/translation="MILLVPLQAQAQVVDNFKLGNVDVSSFKOGLNNRRTKTEK
DPSFSIIQKLEGIPAPKVFVGFISFVGVIVVMYALITKNGQWKNWATGSMI
FESFMRLSRVSLVFLFSDVFNFIPTDFLDLEKATALIFTPFMIAGIRLRRIHE
STKQPYVYRWANRVITGSCLLMLAVLAPWLFANN"
complement(4159. .4788)
/gene="BXA0008"
complement(4159. .4788)
/gene="BXA0008"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX01-04)"
/protein_id="AAM25965.1"
/db_xref="GI:20520083"
/translation="MNISSNNVVEKLEQVRYVEKQVQVLRVSKSSNHHAYMIFFP
EFSVKISNTELENSTVGVYIDRKLQSQEIKEKHNSFIIMTCEVEDVIDDE
TIFNTYVPTTIANKVFVGLGNRLSKLELHTVILNNINRIEIEFVFNENET
VAVEKEIANATMGAYTKANIYYAADLYAAEGNEIKLKSIPKLAGDV"
complement(5124. .5837)
/gene="BXA0009"
complement(5124. .5837)
/gene="BXA0009"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX01-05)"
/protein_id="AAM25966.1"
/db_xref="GI:20520084"
/translation="MFTKPIGIFGVKVFIVVNCSSKKFLREBKKKEMKNPLRPVY
WCPENRPIGRQERTVTPKTEIKDVASIDLVATISVHKNLGEVSRDLRKHDE
IEVYAEVDHPTDPAAVDYIHFPSDEDIPVLELLEDYMPQOEYVSIPTETVPFVQ
QRVADVKMDQFVRWTKVYLBQEQGLTVIRNNNNKRYWIHTBSGELIPNQLLYKLEY
DSPSYTLVWENRQIGNI"
complement(5854. .6561)
/gene="BXA0010"
complement(5854. .6561)

```

Db 111464 GCAGTCTATTGTTAATACAAATTAAGCGTTTGTAGCTCAGATTATCCAAATTCATGAA 111523  
 QY 98 IletyValValAspAspGlySerAlaAsp-----GluthrGlyIleIysArg 113  
 Db 111524 ATTTCCTTTGTTGATGATGCTAGTAAGGATAAATCGGCTTATGAAGTAGCACTTAAATG 111583  
 QY 114 IletyValValAspAspThrGlyAspLeuSerSer----- 126  
 Db 111584 AGGAGGAACCTCTTCACTACCTCAGGAGAAATTCGCTACCACTAAGAAATATTTGTTCT 111643  
 QY 127 -----AsnValIleValHisArgSerGluIysAsnGlnGlyIysArg 140  
 Db 111644 GAAATATTAGTATTCTCCTGACTTATCGTACATCGTTTACCTAAGATTGCGGGAAGA 111703  
 QY 141 HisAlaGlnAlaTrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160  
 Db 111704 CATGCTCAATATGGCTTTTAAACGGCAACACAGCATGCTATTGTATCCATTGATTCATCA 111763  
 QY 161 AspThrTyrlleTyrlleProAspAlaLeuGluGluLeuLeuIysThrPheAsnAspProThr 180  
 Db 111764 GATGCTGATTGTTGTCCTCAATGCTGTAGAGAGTTATTCGAACCTTTAATGATGAAAA 111823  
 QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200  
 Db 111824 GTAATGGCCACAACTGGTCACGTGACATTCGTAATAGAAATGATAATTTATTAAACAAA 111883  
 QY 201 LeuThrAspIleArgTyrlleAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThr 220  
 Db 111884 CTAATGTATGCTGTTATGACAATGCGTTCGTTGGAGCGTGCACACAGTCGCTGATACA 111943  
 QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrlleArgGluValValPro 240  
 Db 111944 GGAATGTTCTGTTGTAGTGGCGCTTAAGTTGTTATTCGTAGAGAGTAATTAACCTGAA 112003  
 QY 241 AsnIleAspArgTyrlleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260  
 Db 112004 AATTTAGAACATATGGAAGTCAGATGTTCTTGGTGGAGGAGTGCAGTTGGAGATGAT 112063  
 QY 261 ArgCysLeuThrAsnTyrlleAlaThrAspLeuGlyIysThrValTyrlleGlnSerThrAlaIys 280  
 Db 112064 AGATGCTCACTAATATGCTATTGTTGAAGGGAACAAAGTTTATCAATCCATCGCTCGA 112123  
 QY 281 CysIleThrAspValProAspLysMetSerThrTyrlleIysGlnGlnAsnArgTTPAsn 300  
 Db 112124 TGTATTACTGATGCTCCACTACATTAACAACTTCTTAACACGCACTACCTTGGAAAC 112183  
 QY 301 LysSerPhePheArgGlnSerIleIleSerValIysIleIleMetAsnAsnProPheVal 320  
 Db 112184 AAGTCATTTTATAGAGAAAGTTTAAATTCATCTGGCATTTGGTATGAAAAACCAAAATGTT 112243  
 QY 321 AlaLeuThrTrpThrIleLeuGluValSerMetPheMetMetLeuValTyrlleValValAsp 340  
 Db 112244 CTGTTTGGCAAAATTTGCAAAATATCGTTATGGATTATTTGGCGTTTCCCTACTCTTA 112303  
 QY 341 PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheValIle 360  
 Db 112304 AGTATTATCTCAAGGCAAGCATGCTAGGTTAATTTGGCTTTATTTATTTGGCTTAT 112363  
 QY 361 IlePheIleValAlaLeuCysArgAsnIleHisTyrlleMetLeuIysHisProLeuSerPhe 380  
 Db 112364 ATTTCATCTGCTATATGCTAGAAATGATTTTATCTATTAATAACATCCCTCTTTTC 112423  
 QY 381 LeuLeuSerProPheTyrlleGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyrl 400  
 Db 112424 TTACTGGCCCAATTATATGGAATCTCCATGCTATTACACACTATTACCTATACGCTTTAT 112483  
 QY 401 SerLeuPheThrIleArgAsnAlaAspTrpGlyThrArg 413  
 Db 112484 GCTTTACTAATCTAATCTAATGTTGGGGAACAGT 112522  
 RESULT 15  
 AF518732  
 LOCUS AF518732 1893 bp DNA linear SYN 28-OCT-2002

DEFINITION Synthetic construct has operon deletion mutant, partial sequence.  
 ACCESSION AF518732  
 VERSION AF518732.1 GI:24415825  
 KEYWORDS synthetic construct  
 SOURCE synthetic construct  
 ORGANISM artificial sequences.  
 REFERENCE 1 (bases 1 to 1893)  
 AUTHORS Walker, J.A. and Timoney, J.F.  
 TITLE Construction of a stable non-mucoid deletion mutant of the  
 Streptococcus equi Pinnacle vaccine strain  
 JOURNAL Vet. Microbiol. 89 (4), 311-321 (2002)  
 MEDLINE 2272956  
 PUBMED 12383640  
 REFERENCE 2 (bases 1 to 1893)  
 AUTHORS Walker, J.A. and Timoney, J.F.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-JUN-2002) Veterinary Science, University of Kentucky,  
 108 Gluck Equine Research Center, Lexington, KY 40546, USA  
 FEATURES  
 source  
 1..1893 /organism="synthetic construct"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32630"  
 <1..>1893  
 /gene="has operon deletion mutant"  
 /note="derived from Streptococcus equi Pinnacle strain"  
 <1..>423  
 /gene="hasA"  
 /note="hyaluronate synthase"  
 430..1137 /note="derived from vector pTW100"  
 <1138..1356  
 /gene="hasB"  
 /note="UDP-glucose dehydrogenase"  
 1406..>1893  
 /gene="hasC"  
 1406..>1893  
 /gene="hasC"  
 /codon\_start=1  
 /transl\_table=11  
 /product="UDP-glucose pyrophosphorylase"  
 /protein\_id="AA559302.1"  
 /db\_xref="GI:24415826"  
 translation="MKVRKAIIPAAIGLGRFELPATKAKEMLPIDVKPTIQIFVEE  
 AKLAEILVYTCVKAKRSIEDHDFDSNEFLYNLQAKGTLLKLVDETTALNLEFIR  
 QSHPGIGDAVLQAKAFVGNFVVLGDDLDITNPSAKPLTKQLIEDIDCTHASTI  
 AV"  
 BASE COUNT 559 a 427 c 385 g 522 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4, 2e-57 Length: 1893  
 Score: 745.00 Matches: 147  
 Percent Similarity: 88.95% Conservative: 6  
 Best Local Similarity: 85.47% Mismatches: 19  
 Query Match: 34.73% Indels: 0  
 DB: 12 Gaps: 0  
 US-09-469-200D-2 (1-417) x AF518732 (1-1893)  
 QY 159 AspSerAspThrTyrlleTyrlleProAspAlaLeuGluLeuLeuLysThrPheAsnAsp 178  
 Db 1 GACTCAGATACTTATATCTACCTCGTGTAGAGGAGCTGTTAAAGACCTTTAATGAC 60  
 QY 179 ProThrValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeu 198  
 Db 61 CCAACTGTTTTTCTGCGACGCGGTCACCTTAATGTCAGAAATAGACAAACCAATCTCTTA 120  
 QY 199 ThrArgLeuThrAspIleArgTyrlleAspAsnAlaPheGlyValGluArgAlaAlaGlnSer 218  
 Db 121 ACACGCTTGACAGATATTCGCTATGATATGCTTTTGGCGTTGAGCAGCTGCCAATCA 180

QY 219 ValThrGlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValVal 238  
 Db |||||  
 181 GTACGGTAATATCTTGTGCTAGGCCACCTAGCGTTACAGCCGAGTGGT 240  
 QY 239 ValProAsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGly 258  
 Db |||||  
 241 GTCTCTAATATAGACAGATACATCAACACAGACCTTCCTGGGTATTCCTGTAAGTATCGGT 300  
 QY 259 AspAspArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThr 278  
 Db |||||  
 301 GATGACAGGTGCTTGACCACTATGCAACTGATTAGGAAAGACTGTTTATCAATCCACT 360  
 QY 279 AlaLysCysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArg 298  
 Db |||||  
 361 GCTAAAGTATTACAGATGTTCTTGACAGATGCCCTACTTACTTGAGCAGCAAAACCGC 420  
 QY 299 TrpAsnLysSerPhePheArgGluSerIleIleSerValLysLysIleMetAsnAsnPro 318  
 Db |||||  
 421 TGGGATCCACTAGTTCTAGAGCGCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCT 480  
 QY 319 PheValAlaLeuThrPheIleLeuGluValSerMet 330  
 Db |||||  
 481 TTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATG 516

Search completed: October 2, 2003, 14:19:11  
 Job time : 5279 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 09:01:58 ; Search time 4785 Seconds  
(without alignments)  
10721.144 Million cell updates/sec

Title: US-09-469-200D-1

Perfect score: 1254

Sequence: 1 atggaacattataaaact.....cacgtataaaattattataa 1254

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 2886711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rtd.\*  
36: em.htg\_mam.\*  
37: em.htg\_vrt.\*  
38: em.sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1254	100.0	1254	1	AF023876	Streptococcus
2	1254	100.0	1254	6	BD087249	Streptococcus
3	1254	100.0	1254	1	AY173078	Streptococcus
4	1228.4	98.0	7268	1	AF347022	Streptococcus
5	623.2	49.7	3466	1	SUB242946	Streptococcus
6	604.6	48.2	1350	1	STRHASAO	Streptococcus
7	604.6	48.2	1440	1	STRHASA	Streptococcus
8	604.6	48.2	1512	6	AR232201	Streptococcus
9	604.6	48.2	10851	1	AE010123	Streptococcus
10	601.4	48.0	9844	1	AE006637	Streptococcus
11	598.2	47.7	53896	1	AE014172	Streptococcus
12	598.2	47.7	323825	1	AF005146	Streptococcus
13	409.2	32.6	1893	12	AF518732	Synthetic
14	282.8	22.6	181654	1	AF065404	Bacillus
15	282.8	22.6	181677	1	AE011190	Bacillus
16	99.4	7.9	6544	1	AF284858	Bradyrhiz
17	73.2	5.8	6014	1	RGNOB	R. gallegae n
18	72.2	5.8	33038	1	AF222753	Bradyrhiz
19	71.8	5.7	110000	1	REU80928.0	Rhizobium e
20	70.2	5.6	230573	1	AF32201252	Bradyrhiz
21	70.2	5.6	299600	1	AF005942	Bradyrhiz
22	70	5.6	813	1	AF217269	Sinorhizo
23	68.8	5.5	843	1	AF217262	Rhizobium
24	68.8	5.5	845	1	AF217263	Rhizobium
25	68.8	5.5	858	1	AF217268	Rhizobium
26	67.2	5.4	871	1	AF217265	Rhizobium
27	66.8	5.3	808	1	AF217264	Rhizobium
28	66.6	5.3	4317	1	ACNOD1	Xi7536 A. caulidans
29	66.6	5.3	9549	1	RHANODABC	L1897 Azorhizobiu
30	66.6	5.3	152050	1	ML0672113	Mesorhizo
31	66.6	5.3	346820	1	AP003008	Mesorhizo
32	65.8	5.2	933	1	AF217266	Rhizobium
33	65.4	5.2	825	1	AF217266	Rhizobium
34	65.4	5.2	839	1	AY166843	Rhizobium
35	65.4	5.2	839	1	AY166846	Rhizobium
36	65.4	5.2	839	1	AY166847	Rhizobium
37	65.4	5.2	900	1	AY166848	Rhizobium
38	65.4	5.2	3403	1	RFNODABCD	X98514 R. tropici n
39	65.2	5.2	894	1	MCI457923	Mesorhizo
40	65.2	5.2	894	1	MME457924	Mesorhizo
41	65.2	5.2	894	1	MME457928	Mesorhizo
42	65.2	5.2	894	1	SSP457926	Sinorhizo
43	65.2	5.2	894	1	SSP457927	Sinorhizo
44	65.2	5.2	934	1	AF217271	Rhizobium
45	65	5.2	2627	1	RLNODABC	X01650 Rhizobium l

## ALIGNMENTS

RESULT 1  
AF023876 1254 bp DNA linear BCT 15-OCT-2002  
LOCUS Streptococcus equisimilis hyaluronan synthase gene, complete cds.  
DEFINITION AF023876  
ACCESSION AF023876  
VERSION AF023876.1 GI:2655099  
KEYWORDS Streptococcus dysgalactiae subsp. equisimilis  
SOURCE Streptococcus dysgalactiae subsp. equisimilis  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.  
REFERENCE 1 (bases 1 to 1254)  
AUTHORS Kumari, K. and Weigel, P. H.  
TITLE Molecular cloning, expression, and characterization of the authentic hyaluronan synthase from group C Streptococcus

equisimilis  
 J. Biol. Chem. 272 (51), 32539-32546 (1997)  
 MEDLINE 98070433  
 PUBMED 9405467  
 2 (bases 1 to 1254)  
 Weigel, P.H. and Kumari, K.  
 Direct Submission  
 TITLE Submitted (10-SEP-1997) Biochemistry & Molecular Biology,  
 JOURNAL University of Oklahoma Health Science Center, P.O. Box 26901, BNSB  
 860, Oklahoma City, OK 73190, USA  
 FEATURES location/Qualifiers  
 source 1..1254  
 /organism="Streptococcus dysgalactiae subsp. equisimilis"  
 /mol\_type="genomic DNA"  
 /sub\_species="equisimilis"  
 /db\_xref="taxon:119602"  
 1..1254  
 /codon\_start=1  
 /transl\_table=11  
 /product="hyaluronan synthase"  
 /protein\_id="AAB87874.1"  
 /db\_xref="GI:2655100"  
 /translation="MRTLKMLTVVAFSPWLLIYVNVLEKAGSLIYGLLIAY  
 LVMSLFFKPFGRAGQKVAALIFSYNEDRESLETLKSVQOQYPLAELIYVD  
 DGSADETGKRIEDYVRDGTGLSSNVVHRSEKQGRHAQWAFERSDADVELTYDS  
 DTYIPDALELLKTFNDVPAATGHLNVRQTNLLTLDIRDNAGVERAAQS  
 VTGNILVSGPLSVREVVPVNDIRYINOTFLGIPVSGIDDRCLNTYATDLGKTVYQ  
 STAKCTIDYDPMKSTYLLKQONRNKSPFRESIISVKKIMNPFVALWTILEVSNFVNL  
 VYSVDFVGNVREDFWLRVLAFLVLIIVALECRNIHYMLKHLPLSFLSPYGVLFH  
 VLQPLKLSLFIIRNADWGRKKLL"  
 BASE COUNT 335 a 240 c 261 g 418 t  
 ORIGIN  
 Query Match 100.0%; Score 1254; DB 1; Length 1254;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-308;  
 Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 ATGAGACATTAAAAACCTCATACCTGTTGGGCTTTAGTATTTTGGGTACTGTTG 60  
 1 ATGAGAACATTAAAAACCTCATACCTGTTGGGCTTTAGTATTTTGGGTACTGTTG 60  
 61 ATTACGTCATCTTATCTCTTTGGTCTAAGGAGCTGTCAATTTATGCTTTTGG 120  
 61 ATTACGTCATCTTATCTCTTTGGTCTAAGGAGCTGTCAATTTATGCTTTTGG 120  
 121 CAGTAGCTACCTATTAGTCAAAATGCTCTTATCTCTTTTACAAGCCATTAAAGGA 180  
 121 CTGATAGCTTACCTATTAGTCAAAATGCTCTTATCTCTTTTACAAGCCATTAAAGGA 180  
 181 AGGCTGGGCAATATAGGTGCGAGCCATTATTCCTCTTATACGAGAGCTGAGTCA 240  
 181 AGGCTGGGCAATATAGGTGCGAGCCATTATTCCTCTTATACGAGAGCTGAGTCA 240  
 241 TTGCTAGAGACTTAAAAAGTGTTCAGCAGCAAACTATCCCTTAGCAGAAATTTATGTT 300  
 241 TTGCTAGAGACTTAAAAAGTGTTCAGCAGCAAACTATCCCTTAGCAGAAATTTATGTT 300  
 301 GTTGACAGTGAAGTCTGATGAGACAGAGTATTAAAGCGCATTAAGACTATGCGCTGAC 360  
 301 GTTGACAGTGAAGTCTGATGAGACAGAGTATTAAAGCGCATTAAGACTATGCGCTGAC 360  
 361 ACTGTGACCTATCAGCAATGCTATTGTTTCATCGGTGAGAGAAATCAAGGAAGGCT 420  
 361 ACTGTGACCTATCAGCAATGCTATTGTTTCATCGGTGAGAGAAATCAAGGAAGGCT 420  
 421 CATGACAGGCTGGGCTTTGAAAGATCAGAGCTGATGCTTTTTTGACCGTGACTCA 480  
 421 CATGACAGGCTGGGCTTTGAAAGATCAGAGCTGATGCTTTTTTGACCGTGACTCA 480  
 481 GATACTTATATACCTGATGCTTTTAGAGAGTGTGTTAAACCTTTTATGACCCAACT 540  
 481 GATACTTATATACCTGATGCTTTTAGAGAGTGTGTTAAACCTTTTATGACCCAACT 540

541 GTTTTGTCTGGCGGGTCACTTAATGTCAGAAATAGACAAACAATCTCTTAACAGC 600  
 541 GTTTTGTCTGGCGGGTCACTTAATGTCAGAAATAGACAAACAATCTCTTAACAGC 600  
 601 TTGACAGATATTCCTATGATATCTTTTGGGCTTGAACAGAGTGCCTTAACTCGTTACA 660  
 601 TTGACAGATATTCCTATGATATCTTTTGGGCTTGAACAGAGTGCCTTAACTCGTTACA 660  
 661 GGTATATATCTTTGCTCAGTCCGCTTGAAGCTTTTACAGACGCGAGTGTGTTCTCT 720  
 661 GGTATATATCTTTGCTCAGTCCGCTTGAAGCTTTTACAGACGCGAGTGTGTTCTCT 720  
 721 AACATAGATAGATACATCAACAGACCTTCTGGGTATTTCCTGTAAGTATTGGTGATGAC 780  
 721 AACATAGATAGATACATCAACAGACCTTCTGGGTATTTCCTGTAAGTATTGGTGATGAC 780  
 781 AGGTGCTTGACCAACTATGCAACTGATTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 840  
 781 AGGTGCTTGACCAACTATGCAACTGATTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 840  
 841 TGTATTACAGATGTTCTCTGACAAGATGCTACTTCTGAAGCAGCAAAACCGCTGGAAC 900  
 841 TGTATTACAGATGTTCTCTGACAAGATGCTACTTCTGAAGCAGCAAAACCGCTGGAAC 900  
 901 AGTCTCTTTTAGAGAGTCCATTAATTTCTGTTAAGAAAATCATGAACAATCCTTTTGTGA 960  
 901 AGTCTCTTTTAGAGAGTCCATTAATTTCTGTTAAGAAAATCATGAACAATCCTTTTGTGA 960  
 961 GCCTATGAGCACTATGAGGTGCTATGTTTATGATGCTGTTTATCTGTGTTGAT 1020  
 961 GCCTATGAGCACTATGAGGTGCTATGTTTATGATGCTGTTTATCTGTGTTGAT 1020  
 1021 TTCTTTTAGGCAATGTCAAGAAATTTGATTGGCTCAGGCTTTTAGCCTTTCGTGTGATT 1080  
 1021 TTCTTTTAGGCAATGTCAAGAAATTTGATTGGCTCAGGCTTTTAGCCTTTCGTGTGATT 1080  
 1081 ATCTCATTTGTCCTGTCGGAACATTCATTACATGCTTAAGACCCGCTGTCCTTC 1140  
 1081 ATCTCATTTGTCCTGTCGGAACATTCATTACATGCTTAAGACCCGCTGTCCTTC 1140  
 1141 TTCTTATCCGTTTATGAGGTGCTGATGTTGTTGCTACAGCCTTGAATATATAT 1200  
 1141 TTCTTATCCGTTTATGAGGTGCTGATGTTGTTGCTACAGCCTTGAATATATAT 1200  
 1201 TCCTTTTACTATTAGAAATGCTGACTGGGGAACAGTAAAAAATTTATTATAA 1254  
 1201 TCCTTTTACTATTAGAAATGCTGACTGGGGAACAGTAAAAAATTTATTATAA 1254  
 RESULT 2  
 BD087249  
 LOCUS Hyaluronan synthase gene and utilization thereof. PAT 27-AUG-2002  
 DEFINITION  
 BD087249  
 ACCESSION  
 VERSION BD087249.1 GI:26532859  
 KEYWORDS JP 2001521741-A/1.  
 SOURCE Streptococcus dysgalactiae subsp. equisimilis  
 ORGANISM Streptococcus dysgalactiae subsp. equisimilis  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 REFERENCE 1 (bases 1 to 1254)  
 AUTHORS Weigel, P.H., Kumari, K. and Deangelis, P.  
 TITLE Hyaluronan synthase gene and utilization thereof  
 JOURNAL Patent: JP 2001521741-A 1 13-NOV-2001;  
 THE BOARD OF REGENTS OF THE UNIVERSITY OF OKLAHOMA  
 COMMENT OS Streptococcus equisimilis  
 PN JP 2001521741-A/1  
 PD 13-NOV-2001  
 PF 30-OCT-1998 JP 200519083  
 PR 31-OCT-1997 US 60/064435; 26-OCT-1998 US 09/178851 PI  
 PAUL H WEIGEL, KSHAMA KUMARI, PAUL DEANGELIS  
 PC C12N15/09, A61K31/728, A61P43/00, C12N9/10, C12P19/04, PC  
 C12Q1/68//

PC (C12N15/09,C12R1:46),C12N15/00,(C12N15/00,C12R1:46) CC  
Hyaluronan synthase gene and utilization thereof FH Key  
Location/Qualifiers  
FT source 1..1254  
FT /organism="Streptococcus equisimilis".  
FEATURES  
source Location/Qualifiers  
1..1254  
/organism="Streptococcus dysgalactiae subsp. equisimilis"  
/mol\_type="genomic DNA"  
/sub\_species="subsp. equisimilis"  
/db\_xref="taxon:119602"  
BASE COUNT 335 a 240 c 261 g 418 t  
ORIGIN  
Query Match 100.0%; Score 1254; DB 6; Length 1254;  
Best Local Similarity 100.0%; Pred. No. 1.5e-308;  
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAGAACATTAAGAACCTCAATACCTGTTGGTCTTATGCTTATTTTGGTACTGTTG 60  
DB 1 ATGAGAACATTAAGAACCTCAATACCTGTTGGTCTTATGCTTATTTTGGTACTGTTG 60  
QY 61 ATTACGTCATGTTATCTCTTTGGTCTTAAAGAGCTTGTCAATTTATGCTTTTGG 120  
DB 61 ATTACGTCATGTTATCTCTTTGGTCTTAAAGAGCTTGTCAATTTATGCTTTTGG 120  
QY 121 CTGATAGCTTACCTATTAGTCAAAATGTCCTTATCCCTTTTAAAGCCATTTAAGGGA 180  
DB 121 CTGATAGCTTACCTATTAGTCAAAATGTCCTTATCCCTTTTAAAGCCATTTAAGGGA 180  
QY 181 AGGGTGGGCAATTAAGTGTGACCAATATTCCCTCTTATACGAAGATCGTAGTCA 240  
DB 181 AGGGTGGGCAATTAAGTGTGACCAATATTCCCTCTTATACGAAGATCGTAGTCA 240  
QY 241 TTGCTAGAGACCTTAAAGAGTGTTCAGAGCAAACTATCCCTTAGCAGAAATTTATGTT 300  
DB 241 TTGCTAGAGACCTTAAAGAGTGTTCAGAGCAAACTATCCCTTAGCAGAAATTTATGTT 300  
QY 301 GTTGAGATGGAAGTGTGATGAGACAGGTATTAAAGCATTTAGAGCTATGCGTGAC 360  
DB 301 GTTGAGATGGAAGTGTGATGAGACAGGTATTAAAGCATTTAGAGCTATGCGTGAC 360  
QY 361 ACTGTGACCTATCAAGCAATGTCTATTCCTGCTCAGAGAGAAATCAAGAAAGCGT 420  
DB 361 ACTGTGACCTATCAAGCAATGTCTATTCCTGCTCAGAGAGAAATCAAGAAAGCGT 420  
QY 421 CATGACAGGCGTGGGCTTTGAAGATCAGACGCTGATGCTTTTGAACGCTTGACTCA 480  
DB 421 CATGACAGGCGTGGGCTTTGAAGATCAGACGCTGATGCTTTTGAACGCTTGACTCA 480  
QY 481 GATACCTATATCTACCTGATGCTTTAGAGAGTGTGTTAAACCTTTAATGACCAACT 540  
DB 481 GATACCTATATCTACCTGATGCTTTAGAGAGTGTGTTAAACCTTTAATGACCAACT 540  
QY 541 GTTTTGTGCGAGGCTACCTTAATGTCAGAAATAGACAAACCAATCTCTTAAACGCG 600  
DB 541 GTTTTGTGCGAGGCTACCTTAATGTCAGAAATAGACAAACCAATCTCTTAAACGCG 600  
QY 601 TTGACAGATATTCGCTATGATATGCTTTGGGCTTGAAGAGCTGCCCATCGGTACA 660  
DB 601 TTGACAGATATTCGCTATGATATGCTTTGGGCTTGAAGAGCTGCCCATCGGTACA 660  
QY 661 GGTAATATCTTTGCTCAGCTCGCTTACGCTTTACAGAGCGAGGTGTTGCTCT 720  
DB 661 GGTAATATCTTTGCTCAGCTCGCTTACGCTTTACAGAGCGAGGTGTTGCTCT 720  
QY 721 AACATAGATAGATACCAACGACCTTCTCGGTATTCCTGTAAGTATGGTGATGAC 780  
DB 721 AACATAGATAGATACCAACGACCTTCTCGGTATTCCTGTAAGTATGGTGATGAC 780  
QY 781 AGGTGCTTGACCAACTATGCAACTGATTTAGGAAGAGCTGTTATCAATCCACTGCTAAA 840  
DB 781 AGGTGCTTGACCAACTATGCAACTGATTTAGGAAGAGCTGTTATCAATCCACTGCTAAA 840

QY 841 TGTATTACAGATGTTCTCTGACAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAAC 900  
DB 841 TGTATTACAGATGTTCTCTGACAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAAC 900  
QY 901 AAGTCCTCTTTTAGAGAGTCCATTATTTCTTTAAGAAATCATGAACAATCCCTTTGTA 960  
DB 901 AAGTCCTCTTTTAGAGAGTCCATTATTTCTTTAAGAAATCATGAACAATCCCTTTGTA 960  
QY 961 GCCCTATGACCACTACTTGAAGTCTATGTTTATGATGCTGTTTATCTGTTGGTGGAT 1020  
DB 961 GCCCTATGACCACTACTTGAAGTCTATGTTTATGATGCTGTTTATCTGTTGGTGGAT 1020  
QY 1021 TTCTTTTAGGCAATGTCAGAGAAATTTGATTTGGCTCAGGCTTTTACGCTTTCTGTTGATT 1080  
DB 1021 TTCTTTTAGGCAATGTCAGAGAAATTTGATTTGGCTCAGGCTTTTACGCTTTCTGTTGATT 1080  
QY 1081 ATCTTCATTTGTCCTGTCGGAACATTCATTACATGCTTAAGCACCCTGCTGCTTC 1140  
DB 1081 ATCTTCATTTGTCCTGTCGGAACATTCATTACATGCTTAAGCACCCTGCTGCTTC 1140  
QY 1141 TTGTTATCTCCCTTTTATGAGGCTGCTGCAATTTGTTCTTACAGCCCTTGAATATAT 1200  
DB 1141 TTGTTATCTCCCTTTTATGAGGCTGCTGCAATTTGTTCTTACAGCCCTTGAATATAT 1200  
QY 1201 TCTCTTTTACTATTAGAAATGCTGCTGGAACAGCTAAAAAATTATTATA 1254  
DB 1201 TCTCTTTTACTATTAGAAATGCTGCTGGAACAGCTAAAAAATTATTATA 1254  
RESULT 3  
AY173078 3552 bp DNA linear BCT 15-JAN-2003  
LOCUS Streptococcus equi subsp. zooepidemicus hyaluronan acid operon,  
DEFINITION Partial sequence.  
ACCESSION AY173078  
VERSION AY173078.1 GI:27762599  
KEYWORDS  
SOURCE Streptococcus equi subsp. zooepidemicus  
ORGANISM Streptococcus equi subsp. zooepidemicus  
REFERENCE 1 (bases 1 to 3552)  
AUTHORS Wu,X., Gao,H. and Chen,J.  
TITLE Molecular characterization of hyaluronan acid operon from group C  
JOURNAL Streptococcus equi subsp. zooepidemicus  
REFERENCE 2 (bases 1 to 3552)  
AUTHORS Wu,X., Gao,H. and Chen,J.  
TITLE Molecular cloning, expression, and characterization of hasC and  
galU, two genes encoding UDP-glucose pyrophosphorylase in group C  
JOURNAL Streptococcus equi subsp. zooepidemicus  
REFERENCE 3 (bases 1 to 3552)  
AUTHORS Wu,X. and Gao,H.  
TITLE Direct Submission  
JOURNAL Submitted (05-NOV-2002) Department of Biology, Tsinghua University,  
Beijing 100084, China  
FEATURES  
source Location/Qualifiers  
1..3552  
/organism="Streptococcus equi subsp. zooepidemicus"  
/mol\_type="genomic DNA"  
/sub\_species="zooepidemicus"  
/db\_xref="taxon:40041"  
1..>3552  
/gene="hyaluronan acid operon"  
1..1254  
/gene="hasA"  
1..1254  
/gene="hasA"  
/notes="required for hyaluronan acid synthesis"  
/codon\_start=1  
/transl\_table=11



## REFERENCE

AUTHORS Blank,L.M., Hugenholz,P. and Nielsen,L.K.  
TITLE Cloning and characterization of the hyaluronic acid synthesis (has) operon from streptococcus equi subsp. zooepidemicus

## JOURNAL

AUTHORS Unpublished

## REFERENCE

AUTHORS Black,L.M. and Nielsen,L.K.  
TITLE Direct Submission  
JOURNAL Submitted (09-FEB-2001) Chemical Engineering Department, The University of Queensland, Coopers Road, Building #74, Brisbane, Queensland 4072, Australia

## FEATURES

## source

1. 7268  
/organism="Streptococcus equi subsp. zooepidemicus"

/mol\_type="genomic DNA"

/sub\_species="zooepidemicus"

/db\_xref="taxon:40041"

469. .>7268

/gene="has operon"

469. .474

/gene="has operon"

481. .>7268

/gene="has operon"

/evidence-experimental

512. 1784

/gene="has"

512. .517

/gene="has"

531. 1784

/gene="has"

/codon\_start=1

/transl\_table=11

/product="hyaluronic acid synthase"

/protein\_id="AAN65248.1"

/db\_xref="GI:24940622"

/translation="MRTLNKLTIVAFSIFVLLIYVNVYLFQAGKSLISYGFLLIAY  
LLVMSLFFYFPFKRAGQYKVAIIPSYNEDAESLLETLKSVQQYPLAEIYVD  
DGADEFGKRTEDYVDPDTGSLSNVIVHRSKQKGRHQAQWAFSDADVETVDS  
DTYVDPALPELLKTFNDPTVFAATGHLAVNRQTNLLTLDIRYDNAGVERAAS  
VTGILVCSPLSYRREYVVPNDIKYINQTPGLIPVSGDRCILNATDLGTVYQ  
STACITDPDKMSTLYLKQQRNKNKFFRESIISVKIMNNPFVALTWLILEVSMFML  
VSYVDFVDNREFDNLRLVLAELVIFITVALCRNIHYMLKPLSLPFPFVGLHLF  
VLQPLKLSLFTIRNADWGRKKLL"

gene

2023. 3242

/gene="hasB"

2023. 2028

/gene="hasB"

2037. 3242

/gene="hasB"

/codon\_start=1

/evidence-experimental

/transl\_table=11

/product="UDP-glucose dehydrogenase"

/protein\_id="AAN65249.1"

/db\_xref="GI:24940623"

/translation="MKISVAGSYVGLSILLSIQAHDVTVVDIDKRVLLINQGISP  
IKDIDAEVLPKNAFLNLTATLDGASVSNADLIITATPNYDSERNVFTDRIIVEVIE  
QVLDINASATIIITKSTPIPLGFKHVREKYQTDRIIFSPFLRSKALYDNLVPSRIIV  
SYEKDSPVIOAKAFAGLLKGAASKDTPVLMGSOBAEVKLEANTFLAMVSYF  
NELDTYISGKGLQKRVIEGVCHQDRIGHNNPSPFGYGCGLPKDKQLLANYRGP  
OSLSMAVESKNIRKSYLAELDRASSQKQAGVPLTIGFYRLIMKNSDNRSAIK  
DTIIDINDYGVNIVIPMLGEDIQYRVKDLQEQFKNESTIIIVSNRFDOLGVDIVDK  
YTRDVFGRD"

gene

3280. 4194

/gene="hasC"

3280. 3285

/gene="hasC"

3292. 4194

/gene="hasC"

/codon\_start=1

/evidence-experimental

/transl\_table=11

/product="UDP-glucose pyrophosphorylase"

/protein\_id="AAN65250.1"

/db\_xref="GI:24940624"

/translation="MTKVKALILPAAGLGRFLPATKALAKEMLPIDVPTQFIYVEE  
ALKAGIBEIIVTVTKAKRSIEDHFDNFLEYNLQAKGTELLKLVDETTAINLHFIR  
QSHPRGLGDAVLAQAKAFVGNFPFVYMLGDDLDITNPSAKPLAKOLIEDYDCTHAST  
AVMRPHEVSNYGVIAQGVKAVKGLYSVETFEKPSPEAPSLAIIGRIYLLPEIF  
AILENOAPGAGNEVOLADAIDKLNKTQRFAREFKGERYDVGDFGFMKTSLSLYALKH  
POYRDDLDYIIILSKQLNKDVKK"

gene

4349. 5744

/gene="glmU"

4349. 4354

/gene="glmU"

4362. 5744

/gene="glmU"

/codon\_start=1

/transl\_table=11

/product="UDP-N-acetyl-glucosamine pyrophosphorylase"

/protein\_id="AAN65251.1"

/db\_xref="GI:24940625"

/translation="MKNVAILAAGKTRMNSGLSKVLHKVSGLSMLEHVLKVSALA  
POKQLTVIGHQAOQVRAVLGOLLTVYVQEQLGTGSHAVMAEBELSLGQLVLAGD  
TPLIRGSLKALDIYHIREKNVATILITANAKDPFGRLIRNAAGEVYVNIIVSQKDANE  
AEQVEKINTGTVIFDNKELFALKHLITDNAQGEYITDVISFASKEKYGAYLLK  
DEDESLGYNDRALAQAEVIMQERINKQHMLGTQLQNPAAATYTESSEIAPDVLIEA  
NVTLGQTRIGRSRVITNGSYILDSRLGEGVYVSQSVLEGVADYVTPYGAHTRPD  
SOLDECVHIGNEVEKSHLGANTKAGHLTYLIGNAEIGSEVNTGAGSITVNYDQGRY  
QTVIGDHAFITGSHSTLIAPVEGENALTAAGSTIAGSVPAUSVAIGRSQVYKEGYAK  
RUPHPDQDQ"

gene

5905. 7268

/gene="pgi"

5905. 5910

/gene="pgi"

5919. 7268

/gene="pgi"

/codon\_start=1

/transl\_table=11

/product="glucose-6-phosphate isomerase"

/protein\_id="AAN65252.1"

/db\_xref="GI:24940626"

/translation="MSHIFDYXKVLQFAGQHEIDFLOGOVTEADQALRQGTGPGSD  
FLWLELPENDKEERARILKAKELIKADSDVLVIGIGSGVLGAKAKAIDFLNSHPAN  
LOTAKERAPDIIYAGNLSISSYLDLVYQDKDFSNVLSKSGTTEPFAFAFRFK  
ELLVKKYQEEANKRIYATDKVGAKEADANHWETVVDNVGNVFRFSVTAFLVLL  
PIAAGADITALEMAGANAARKLSSDKISENIAYQAVVYVNLIRKGYVTEILANYPE  
SLQYSENNKOLAGESBQKQIYPTSANFSLDLHSLGQFOEGYRNLPETVRVDK  
PGONVILPEMAEDLDGLYLOQKVDYFNKATDGVLLAHTDGGVNMFTLPEODEF  
TLGYIYFFELAIHLSYINGVNPDPQVGEAKYKNMFALLGKQFQFELGAALNRL"

BASE COUNT

2105 a 1485 c 1733 g 1945 t

ORIGIN

Query Match 98.0%; Score 1228.4; DB 1; Length 7268;

Best Local Similarity 98.7%; Pred. No. 5.4e-302;

Matches 1238; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 ATGAGACATATAAAACCTCATACCTGTGTGGCTTTAGTATTTTTTGGGTACTGTTG 60

Db 531 ATGAGACATATAAAACCTCATACCTGTGTGGCTTTAGTATTTTTTGGGTACTGTTG 590

Qy 61 ATTTACCTCAAGCTTTACTCTTTGGTGGCTTAAGCAAGCTTGTCATTTATGGCTTTTG 120

Db 591 ATTTACCTCAAGCTTTACTCTTTGGTGGCTTAAGCAAGCTTGTCATTTATGGCTTTTG 650

Qy 121 CTGATAGCTTACCTATTAGTCAAAATGCTCTTATCCTTTTTTACAAGCATTAAAGGA 180

Db 651 CTGATAGCTTACCTATTAGTCAAAATGCTCTTATCCTTTTTTACAAGCATTAAAGGA 710

Qy 181 AGGGCTGGCAATATAAGGTTGACGACCATTTATCCCTCTTATTAACGAAGTCTGAGTCA 240

Db 711 AGGGCTGGCAATATAAGGTTGACGACCATTTATCCCTCTTATTAACGAAGTCTGAGTCA 770

Qy 241 TTGCTAGAGACCTTAAAGGTTTCAGCAGCAACCTATCCCTTACGAGAAATTTATGTT 300

Db 771 TTGCTAGAGACCTTAAAGGTTTCAGCAGCAACCTATCCCTTACGAGAAATTTATGTT 830

```
QY 301 GTTACGATGGAAGTCTGATGACAGAGTATTAGCGCATTTAGCACTATGTCGGTAC 360
Db 831 GTTACGATGGAAGTCTGATGACAGAGTATTAGCGCATTTAGCACTATGTCGGTAC 890
QY 361 ACTGCTGACCTATCAAGCAATGTCATTTTCATCGTTCAGAGAAAATCAAGAAAGCGT 420
Db 891 ACTGCTGACCTATCAAGCAATGTCATTTTCATCGTTCAGAGAAAATCAAGAAAGCGT 950
QY 421 CATGACAGCCCTGGCCCTTTGAAGATCAGACGCTGATGCTTTTTCAGCCGTGACTCA 480
Db 951 CATGACAGCCCTGGCCCTTTGAAGATCAGACGCTGATGCTTTTTCAGCCGTGACTCA 1010
QY 481 GATACCTATATCTACCTGATGCTTTAGAGAGTGTGTAAACACCTTTAATGACCAACT 540
Db 1011 GATACCTATATCTACCTGATGCTTTAGAGAGTGTGTAAACACCTTTAATGACCAACT 1070
QY 541 GTTTTTCGTCGACGGTCACTTAATGTCAGAAATAGACAAACCAATCTTTAACAGC 600
Db 1071 GTTTTTCGTCGACGGTCACTTAATGTCAGAAATAGACAAACCAATCTTTAACAGC 1130
QY 601 TTGACAGATATCGCTATCAATGCTTTTGGCGTTGACGCTGCCAATCGTTACA 660
Db 1131 TTGACAGATATCGCTATCAATGCTTTTGGCGTTGACGCTGCCAATCGTTACA 1190
QY 661 GGTATATCTCTTTGCTCAGGTCGCTTTAGCGTTTACAGACGAGTGGTTGCTTCT 720
Db 1191 GGTATATCTCTTTGCTCAGGTCGCTTTAGCGTTTACAGACGAGTGGTTGCTTCT 1250
QY 721 ACATAGATAGATACATCAACAGACCTTCCTGGGTATCTGTAGTATTTGGTATGAC 780
Db 1251 ACATAGATAGATACATCAACAGACCTTCCTGGGTATCTGTAGTATTTGGTATGAC 1310
QY 781 AGGTGCTTGACCAACTATGCAACTGATTTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 840
Db 1311 AGGTGCTTGACCAACTATGCAACTGATTTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 1370
QY 841 TGTATACAGATGCTCCTGACAGATGCTTACTTACTTGAAGCAGCAAAACCGCTGGAAC 900
Db 1371 TGTATACAGATGCTCCTGACAGATGCTTACTTACTTGAAGCAGCAAAACCGCTGGAAC 1430
QY 901 AGTCTCTCTTTAGAGTCCATTTATCTGTAGAAATCATGACATCTCTTTTCTA 960
Db 1431 AGTCTCTCTTTAGAGTCCATTTATCTGTAGAAATCATGACATCTCTTTTCTA 1490
QY 961 GCCCATGACCACTATGAGTGTCTATGTTATGATGCTGTTTATCTGTGTTGAT 1020
Db 1491 GCCCATGACCACTATGAGTGTCTATGTTATGATGCTGTTTATCTGTGTTGAT 1080
QY 1021 TTTCTGTAGCAATGTCAGAAATTTGATGCTCAGGTTTTCAGCCCTTCTGTTGAT 1140
Db 1551 TTTCTGTAGCAATGTCAGAAATTTGATGCTCAGGTTTTCAGCCCTTCTGTTGAT 1200
QY 1081 ATCTTCATTTGTCCTGTCGGAACATTCATTATGCTTAAGCAGCCGCTGCTCTTC 1140
Db 1611 ATCTTCATTTGTCCTGTCGGAACATTCATTATGCTTAAGCAGCCGCTGCTCTTC 1200
QY 1141 TTGTTATCTCGTTTATGCGGTGCTGATTTGTTGCTTACAGCCCTTGAATATAT 1260
Db 1671 TTGTTATCTCGTTTATGCGGTGCTGATTTGTTGCTTACAGCCCTTGAATATAT 1320
QY 1201 TCTCTTTTACTATAGAAATGCTGCTGGGACACGTAATAAATTTATATAA 1254
Db 1731 TCTCTTTTACTATAGAAATGCTGCTGGGACACGTAATAAATTTATATAA 1784
```

## RESULT 5

```
SUB242946 LOCUS 3466 bp DNA linear BCT 06-JAN-2001
DEFINITION Streptococcus uberis hasa and hasb genes for putative hyaluronan
synthase and UDP-glucose dehydrogenase.
ACCESSION AJ242946
VERSION AJ242946.2 GI:7532762
```

```
KEYWORDS hasa gene; hasB gene; hyaluronan synthase; UDP-glucose
dehydrogenase.
SOURCE Streptococcus uberis
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Ward,P.N., Field,T.R., Ditcham,W.G., Maguin,E. and Leigh,J.A.
Identification and disruption of two discrete loci encoding
hyaluronan acid capsule biosynthesis genes hasa, hasB, and hasC in
Streptococcus uberis
Infect. Immun. 69 (1), 392-399 (2001)
20569205
11119529
REFERENCE
2 Ward,P.N.
Direct Submission
Submitted (07-JUN-1999) Ward P.N., Environmental Microbiology,
Institute for Animal Health, Compton Laboratory, Compton, nr
Newbury, Berkshire, RG20 7NN, UNITED KINGDOM
Revised by [3]
3 (bases 1 to 3466)
Ward,P.N.
Direct Submission
Submitted (07-APR-2000) Ward P.N., Environmental Microbiology,
Institute for Animal Health, Compton Laboratory, Compton, nr
Newbury, Berkshire, RG20 7NN, UNITED KINGDOM
On Apr 10, 2000 this sequence version replaced gi:5441847.
FEATURES
location/Qualifiers
1..3466
/organism="Streptococcus uberis"
/mol_type="genomic DNA"
/strain="0140J"
/db_xref="taxon:1349"
/country="United Kingdom"
23..1276
/gene="hasa"
23..1276
/gene="hasA"
/function="putative biosynthesis of hyaluronic acid"
/codon_start=1
/transl_table=11
/product="hyaluronan synthase"
/protein_id="CAB46918.2"
/db_xref="GI:7532763"
/translation="MEKLNLTPTMTFPLWLIIGLVNFVFGTKGLVYGIILLY
LSIKMGLSFYRYPKGYGVKVAALIPSYNEDGVGLLETKSVOKOTPIAEIFVID
DSVDKTKGLVEDYKLVNGFGDQVIYHQPENVGKRAQAFAEESADIVFLVDS
TIIPDALBELLTAFNDPEVTAANTGHLNARNKOTNLRLDIRDNAGVERAAQSV
TGNILVCGPLSYRRSVGIPNDRYTSQTFGLVPVSGIDDRCLTNYATDLGKTYQS
TARCTDVPDKFVKFIQONRMKSFRESIISVKLLATPSAVMTITEVSMFTMLV
YSIFSLIGBAQGFNLKLVAFVLIIIFIVALCRNHYVMVKHPFAFLSPFYGLIHLEF
LQPKIYSLFTIRNATGTRKTKSK"
1301..2506
/gene="hasB"
1301..2506
/gene="hasb"
/function="putative biosynthesis of D-glucuronic acid"
/codon_start=1
/transl_table=11
/product="putative UDP-glucose dehydrogenase"
/protein_id="CAB46919.2"
/db_xref="GI:7532764"
/translation="MKIIVAGSGVGLSVLLAOKNPVTVDIEKKVNLINOKSP
IQVDIENLYLKEKKLQRLADQAEADAILIATPTNYDVEKNFPTSHVETVIE
KALANSGALLVYKSTPIGLTKMKRKQYQTRIIIFSPFEKRSKALKNLNLPRIIV
SPEDDSSNEVEIAKTFQALLKDSGLDKVPVLFMGSAEAVALFANTYLAHRVYTF
NELDTYAEKGLNDRNIIEGVCHDRIGHYNNPFGYGGYCLPKDTKQLLAGYDGP
QSLIKAIYDSNKIRKEYIASQTLQQLSDINVPDKATIGIYRLIMKNSDNPRESSAIK
DIIDHIKSYQINIVLYPEPMNEDFLPDIIDLSDFKAMSHIYSNRYDLALEDVKEKY
VTRDIYGDV"
```



```

/ gene="hasA"
79..1338
/ gene="hasA"
/ function="required for hyaluronate biosynthesis"
/ note="GTC start codon"
/ codon_start=1
/ transl_except="(pos:79..81,aa:Met)
/ transl_table=11
/ product="hyaluronate synthase"
/ protein_id="AA17725.1"
/ db_xref="GI:473747"
/ translation="MPFKKLLIVSFILISILILYLMYLFSTVIGYVILITL
VILGLSFIYEFKPKDYKVAAPISYNEDAESLETLKSLAQTYPLSEIYVDD
GSSNTAIOLEIEYVREVDICRNVLVNRKGRHQAQAFERSADVFLTVDSL
TYIYPNALSELKSFNDETVYAATHLNARNQNLRLTRIDRYDNAGFVERAOSL
TGNIYVCSPLSIYRVRVILPNLRYKQTFGLPVSIGSDRLCLNYAIDLGPVYQS
TARCDVDVPSLKYLVKQORNKSFRESIISVKKILSNPIVALWTIFEVWPMELI
VATGNLLFNQAIQDLKLFALFSLIFIVALCRNVHVMKHPASELLSPLYGILHLFV
LQPKILISLCTIKNTGWRKKVTIFK"
1002
misc_feature
/ gene="hasA"
/ note="site of Tn916 insertion in WF62 (hyaluronate
synthase mutant)"
BASE COUNT 427 a 217 c 223 g 483 t
ORIGIN
Query Match 48.2%; Score 604.6; DB 1; Length 1350;
Best Local Similarity 68.7%; Pred. No. 4.4e-143;
Matches 848; Conservative 0; Mismatches 384; Indels 3; Gaps 1;
QY 10 TTAAAAACCTCAATACGTGTTGGCCCTTAGTATTTTTGGGTACTGTTGATTTACGTC 69
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 88 TTTAAAAAACTTTAATGTTTATTCCTTTATCTTTTATTTTGGTATATCTATCTATCTA 147
QY 70 AATGTTATCTCTCTGGTCTAAAGGAGCTTGCAATTTATGCTTTTGGTGTAGTACT 129
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 148 AATATGATCTATTGGAC---ATCACTGTAGAAATTAAGAGTAATATTAATACC 204
QY 130 TACCTATAGTCAAAATGCTCTATCCCTTTTACAGCAATTTAAGGAAAGGCTGGG 189
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 205 TATCTAGTTATCAAACTTGGATTATCTTCTTTATGAGCAATTTAAAGAAAGCCAT 264
QY 190 CAATATAGCTGCGACCAATATCCCTCTTATACGAGATGCTGAGTCATGTCAGAG 249
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 265 GACTATAAGCTGCTGTGTAATTCCTTTCTTATATGAGATGCCGAGTCATATAGAA 324
QY 250 ACCTTAAAAAGTGTTCAGCAGCAAACTATCCCTTAGCAGAAATTTATGCTTGACGAT 309
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 325 ACACCTTAAAAGTGTGTAGCAGACACCTATCCGTTATCAGAAATTTATATGTTGATGAT 384
QY 310 GGAAGTCTGATGAGACAGGTATTAAGCCGATTAAGACTATGTCGTCACACTGGTGAC 369
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 385 GGGAGTTCAAACACAGATGCCAATACAATTAATGAAGAGTATGTAATAGAGAAGTGAT 444
QY 370 CTATCAGCAATGTCATGTTTCATCGCTCAGAGAAAATCAAGAAAGCTCAGCCACAG 429
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 445 ATTGTCGAACGTTATGTTTACCGTTCCTTGTCATAAGAAAGAAAGCCATGCTCAA 504
QY 430 GCGTGGCCCTTTGAAGATCAGACGCTGATGCTTTTTTGACCGCTTGACTCAGATACATAT 489
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 505 GCGTGGCACTTTGAAGATCTGACGCTGACGTTTTTTTACCGTAGATTCAGATACATAT 564
QY 490 ATCTACCCCTGATGCTTTAGAGAGTGTGTTAAAAACCTTTAATGACCAACTGTTTTGCT 549
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 565 ATCTATCAAAATGCCCTTAGAAGAACTCTAAAAAGCTTCAATGATGAGACAGTTATGCT 624
QY 550 GCGAGGCTCACCTTAAGTCAGAAATAGACAAACCAATCTCTTACAGCTTGACAGAT 609
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 625 GCACAGACATTTGANTGCTAGAACACAGCAAACTAATCTATTAACGCGACTTACAGAT 684
QY 610 ATTCGCTATGATATGCTTTTGGGTTGACGAGCTGCCAAATCCGTTACAGGTAATATC 669
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 685 ATCCGTTACGATATGCTTTGGGGTGGAGCGTGTGTCATCATTTACAGGTAATATT 744

```

```

QY 670 CTGTTTGTCTAGTCGCTTAGCGTTTACAGACGCGAGGTGGTTGTTTCCTTAACATAGAT 729
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 745 TTAGTTTGTCTAGGACCATTTAGATTTATTCGAGGTGAAGTGAATTTTCCTTAACATAGAG 804
QY 730 AGATACATCAACAGACCTTCTGGGTATTCCTCTAGTATTGTTGTCAGACAGTCTTG 789
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 805 CGGTATAAATCAACATCTCTAGTTTACCTGTAGCATTTGGGGATGATCATGATTTA 864
QY 790 ACCAATATGCAACTGATTTAGGAAGACTGTTTATCAATCCACTGCTAAATGATTACA 849
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 865 ACAATATATGCTATTGATTTAGGACGCACTGCTTACCAATCAACAGCTAGATGATACT 924
QY 850 GATGTTCTGCAAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAACAAGTCCCTC 909
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 925 GATGTACCTTCCCAATTAAGAGTTATTTAAGCAACAAATCGATGGAATAAATCTTTT 984
QY 910 TTTAGAGAGTCCATATTTCTTGAAGAAATCATGAACAATCCTTTTGTAGCCCTATGG 969
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 985 TTTAGAGAATCTATTTCTGTAAATAAATCTTTCTAATCCCATCGTTCCTTATGG 1044
QY 970 ACCATATCTAGGTGCTCTATGTTTATGATGCTGTTTATCTGTGGTGGGATTTCTTTGTA 1029
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1045 ACTATTTTGAAGTCTGTTATGTTGATGTTGATGTCGAATTTGGAAATCTTTTGT 1104
QY 1030 GGCAATGTCAGAGAAATTTGATTTGGCTCAGGGTTTATAGCCCTTCTGTCATATCTTCAT 1089
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1105 AATCAAGCTATTCATTTAGACCTTATTAACCTTTTGGCTTTTATCCATCATCTTTATC 1164
QY 1090 GTTGCCCTGTGTCGAACATCATTTACATGCTTAAGCACCCGCTGCTCTCTTCTTTGTTATC 1149
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1165 GTTGCTTTATGCTGAATGTTTATGTTATGTTCAACATCTCTGCTAGTTTGTGTTATCT 1224
QY 1150 CCGTTTATGCGTCTGATGTTTGTCTCAGACCCCTTGAAATATATTTCTCTTTT 1209
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1225 CTTCTGTAGGAATTTACACTGTTTGTCTACAGCCCTTAAACTTATCTTTATGTC 1284
QY 1210 ACTATTAGAAATGCTGACTGGGGAACAGTAAAAA 1244
DB 11 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1285 ACCATTAATAATACGAATGGGGAACAGTAATAA 1319
STRHA5 1440 bp DNA linear BCT 09-MAY-1994
Streptococcus pyogenes hyaluronan synthase (hasA) gene, complete
cds
ACCESSION L20853
VERSION L20853.1 GI:410042
KEYWORDS hasA gene; hyaluronan synthase; hyaluronate; hyaluronic acid.
SOURCE Streptococcus pyogenes
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 1440)
AUTHORS DeAngelis,P.L., Papaconstantinou,J. and Weigel,P.H.
TITLE Molecular cloning, identification, and sequence of the hyaluronan
synthase gene from group A Streptococcus pyogenes
J. Biol. Chem. 268 (26), 19181-19184 (1993)
MEDLINE 93374890
PUBMED 8366070
COMMENT Original
FEATURES
source
Location/Qualifiers
1..1440
/organism="Streptococcus pyogenes"
/strain="543"
/db_xref="taxon:1314"
109..1380
/gene="hasA"
109..114
/gene="hasA"
/ note="possible ribosome binding site at -109 (upstream of
atg start codon) near alternate start codon (GTG) at

```



Db 1147 ATCAAGTATCAATAGACCTTATTAACCTTTTGGCTTTTATCCATCATCTTTATC 1206  
 QY 1090 GTTGCCCTGTCGGAACATCATTAACATGTAAGCACCCTGCTCTCTCTGTTATCT 1149  
 Db 1207 GTTGCTTTATGTCGTAATGTTTCAATTAATGTCGAACATCTCTGTTATCT 1266  
 QY 1150 CGTTTATGCGGTCGCAATGTTTGTCTACAGCCCTTGAATATATATCTCTTTT 1209  
 Db 1267 CCTGTATGAATATATACATGTTTGTCTACAGCCCTTGAATATATATCTCTTTT 1326  
 QY 1210 ACTATTAGAAATGTCGCTGAGGGAACAGTAAAA 1244  
 Db 1327 ACATTAATAATGGAATGGAACAGTAAAA 1361

RESULT 8  
 AR232201  
 LOCUS 1512 bp DNA linear PAT 20-DEC-2002  
 DEFINITION Sequence 1 from patent US 6455304.  
 ACCESSION AR232201  
 VERSION AR232201.1 GI:27274189  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 1512)  
 AUTHORS Weigel, P.H., DeAngelis, P.L. and Papaconstantinou, J.  
 TITLE Hyaluronate synthase gene and uses thereof  
 JOURNAL Patent: US 6455304-A 1 24-SEP-2002;  
 FEATURES Location/Qualifiers  
 source 1..1512  
 BASE COUNT 484 a 227 c 245 g 556 t  
 ORIGIN

Query Match 48.2%; Score 604.6; DB 6; Length 1512;  
 Best Local Similarity 68.7%; Pred. No. 4.4e-143;  
 Matches 848; Conservative 0; Mismatches 384; Indels 3; Gaps 1;

QY 10 TTAATAACCTCATACTGTTGGCTTTAGTATTTTGGTGACTGTTGTTACGTC 69  
 Db 202 TTTAAAAAATTAATGTTTATCTTTTATCTTTTATCTTTTATCTTTATCTA 261  
 QY 70 AATGTTTATCTTTGCTGCTAAAGGAGCTTCTCAATTTAGCTTTTGTCTGATCT 129  
 Db 262 AATAGTAATCTTTGGAAC---ATCACTGTAGAAATTTATGAGTAATATTAATACC 318  
 QY 130 TACCTATTAGTAAATGTCCTTATPCCTTTTACAAGCCATTTAAGGAAGGCTGGG 189  
 Db 319 TATCTAGTTATCAACTTGGATTATCTTTCTTTATGAGCCATTTAAGGAATCCAT 378  
 QY 190 CAATATAGGTTGCACCAATTTATCCCTCTTATAGGAAGATGCTGAGTATGCTAG 249  
 Db 379 GACTATAAGTGTCTGCTTAATCTTCTTATATGAAGATGCCGAGTCAATTATGAA 438  
 QY 250 ACCTTAAAGTGTTCAGCAGCAAACTATCCCTAGCAGAAATTTATGTTGTCACCAT 309  
 Db 439 ACACCTTAAAGTGTTCAGCAGCACTATCCCTATCAGAAATTTATGTTGATGAT 498  
 QY 310 GGAAGTGTGATGAGACAGGTATTAAGCCATTTAAGCAATGTCGCTGACACTGGTGAC 369  
 Db 499 GGGAGTTCAACACACAGATGCAATTAATTAAGAGTATGTAATAGAGAGTGGAT 558  
 QY 370 CTATCAACATGTCATTTCTTCGCTGAGAAATCAAGAAAGCTCATGACAG 429  
 Db 559 ATTTCGAAACGTTATCTGCTTACCGTTTCCCTTTGCAATAAAGGAACCCATGCTCAA 618  
 QY 430 GCGTGGCTTTGAAAGATCAGACGCTGATGCTTTTTCAGCGTTGACTCAGATCTTAT 489  
 Db 619 GCGTGGCAATTTGAAAGATCTGACGCTGAGCTTTTTCAGCGTACTCAGATCTTAT 678  
 QY 490 ATCTACCTGATGCTTTAGAGGAGTGTGTAATAACCTTTAATGACCAACTGTTTTCCT 549

Db 679 ATCTATCCAAATGCCCTTAGAGAACTCTTAAAGCTTCAATGATGACAGAGTTATGCT 738  
 QY 550 GCGAGGGTCACTTAAATGTCAGAAATAGACAAACCAATCTCTTTAACAGGTTGACAGAT 609  
 Db 739 GCAACAGGACATTTGAATGCTAGAAAACAGACAACTAATCTATTAAACGCACTTACAGAT 798  
 QY 610 ATTGCGTATGATAATGCTTTTGGCTTTGAACGAGTGCCTCAATCCGTTACAGGTAATATC 669  
 Db 799 ATCCGTTAGTAATGCTTTGGGGTGGAGCGTCTCTCAATCATTAACAGGTAATAT 858  
 QY 670 CTGTTTGTCTCAGTCCGCTTAGCTTTTACAGACGAGGTTGTTTCTTAAACATAGAT 729  
 Db 859 TTAGTTTGTCTCAGAGCACTTGAATTTATTCGACGTGAAGTATTTCTTAACTTAGAG 918  
 QY 730 AGATACATCAACAGACCTTCCCTGGGTATCTCTGTAAGTATTTGGTGATGACAGGTCGTTG 789  
 Db 919 CGCTATAAAAATCAAACTTCTAGCTTTTACCTGTTAGCAATTTGGGGATGATCGATGTTA 978  
 QY 790 ACCAATATGCAACTGATTTAGGAAAGACTGTTTATCAATCCACTGCTAAATGTTATTACA 849  
 Db 979 ACAAAATATGCTATTGATTTAGGACGCACTGCTACCAATCAGACGATGATGATGATCT 1038  
 QY 850 GATGTTCTCTGACAGATGCTCTACTTACTTGAAGCAGCAAAACGCTGGACAACTCTTC 909  
 Db 1039 GATGTACCTTTCCAAATTAAGAGTATTTTAAAGCAACAAATCGATGGAATAATCTTTT 1098  
 QY 910 TTTAGAGAGTCCATTTATTTCTGTTAGAAATCATGACAAATCCCTTTTGTAGCCCTATGG 969  
 Db 1099 TTTAGAGATCTATTTTCTGTTTAAATAATCTTTCTTAATCCCATGTTGCTCTATGG 1158  
 QY 970 ACCATACCTTGAGGTGCTGATGTTTATGATGCTTTTATTTCTGCTGATTTCTTTGTA 1029  
 Db 1159 ACTATTTTCGAAGTCGTTATGTTTATGATGTTGATTTCTGCAATTTGGAATCTTTGTT 1218  
 QY 1030 GGCAATGTCAGAGATTTGATTTGGCTCAGGTTTACCTTTTCTGGTGATTAATCTCAT 1089  
 Db 1219 AATCAAGCTATTCAATAGACCTTATTAACCTTTTTCCTTTTATCCATCATCTTTATC 1278  
 QY 1090 GTTGCCCTGTCGCGAATCATTAATGCTTAAGCACCCTGCTGCTCTCTCTGTTATCT 1149  
 Db 1279 GTTGCTTTATGCTGTAATGTTTATTTATGTTGCTCAACATCTGCTAGTTTGTGTTATCT 1338  
 QY 1150 CCGTTTATGCGGTCGCTGCAATTTGTTTGTCTACAGCCCTTGAATATATATCTCTTTT 1209  
 Db 1339 CCTGTATGGAATATATACATTTGTTGCTTACAGCCCTTAAACTTTATCTTTATGCT 1398  
 QY 1210 ACTATTAGAAATGCTGATGCTGGGAACAGTAAAA 1244  
 Db 1399 ACCATTAATAATGGAATGGAACAGTAAAA 1433

RESULT 9  
 AE010123  
 LOCUS 10851 bp DNA linear BCT 03-APR-2002  
 DEFINITION Streptococcus pyogenes strain MGAS8232, section 171 of 173 of the complete genome.  
 ACCESSION AE010123 AE009949  
 VERSION AE010123.1 GI:19749282  
 KEYWORDS  
 SOURCE Streptococcus pyogenes MGAS8232  
 ORGANISM Streptococcus pyogenes MGAS8232  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.  
 REFERENCE 1 (bases 1 to 10851)  
 AUTHORS Smoot, J.C., Barbican, K.D., Van Gompel, J.J., Smoot, L.M., Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M., Porcella, S.F., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M., Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.G. and Musser, J.M.  
 TITLE Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)  
 MEDLINE 21927593

11917108

PUBMED  
REFERENCE  
AUTHORS

2 (bases 1 to 10851)  
Smoot,J.C., Barbican,K.D., Van Gompel,J.J., Smoot,L.M.,  
Chaussee,M.S., Sylva,G.I., Sturdevant,D.E., Ricklefs,S.M.,  
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,  
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.George. and Musser,J.M.  
Direct Submission  
Submitted (31-JAN-2002) Laboratory of Human Bacterial  
Pathogenesis/Rocky Mountain Laboratories/NIH, 903 S. 4th St.,  
Hamilton, MT 59840, USA

TITLE  
JOURNALFEATURES  
source

Location/Qualifiers

1..10851

/organism="Streptococcus pyogenes MGAS8232"

/mol\_type="genomic DNA"

/strain="MGAS8232"

/db\_xref="taxon:186103"

complement(238..852)

/gene="spym18.2226"

complement(238..852)

/gene="spym18.2226"

/note="best blastp match gb|AAK34820.1| (AE006636)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAL98659.1"

/db\_xref="GI:19749283"

/translation="MFKENLKORVENGLVALALTILAIIEAFSKNADPKYAKKS  
ESKWTIDAPKNHATIKESKEKAKSIASEPIPTVENSVAPTVEAPVQOEVTQ  
TVQGVSSVAINFNVLISNGTAGIVGQAQAQAAATVGPSTWEHIIARENGNP  
AANASGSLFQTPMGWGSTATVEDQVNAALKAYSAQGLSANGY"

complement(1223..2023)

/gene="spym18.2228"

complement(1223..2023)

/gene="spym18.2228"

/note="best blastp match gb|AAK34821.1| (AE006636)  
conserved hypothetical protein [Streptococcus pyogenes M1  
GAS]"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAL98660.1"

/db\_xref="GI:19749284"

/translation="MDKLILGRYPGDSLIHRLDPRSKLLMIYIVIFWANNVNT  
DLGISOAILIFNRVLIIFSTLLITLTPLSDAVESLLKPLTRFKVPAHETGLM  
SLSRVPTLMDTTRIMNAQRAGVDREGNLQKVKSIIPIILIPFASSEKADAL  
AIAMEARGVQSGEGRTYRQLDQLKSIAIGVISLGLLILFLKTSI"

complement(2016..2858)

/gene="spym18.2229"

complement(2016..2858)

/gene="spym18.2229"

/note="best blastp match gb|AAK34822.1| (AE006636)  
putative ABC transporter (ATP-binding protein)

[Streptococcus pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="putative ABC transporter (ATP-binding protein)"

/protein\_id="AAL98661.1"

/db\_xref="GI:19749285"

/translation="MSINLQNSVTVQAGTPEGRALFNLDLGDYTAFIGTGS  
GKSTIQMLNGLHVPTGTIVSDQDITNHNKEIKSIKKHVLGVFPFESQLFEET  
VLKDVAFQNGVSPFEALAREKALALVGSINLEFKNPFELSGQVRRAVAGIL  
AMQPKVLVDEPTAGLDPKGRKEIMTIFKHKQSGMTIVLVTHLMDADVADFVYL  
DKGKILSGKPTITFQQVSLLEKKQLGVKVTKLAQRLVDRGIPISLPILEELREV  
LKHC"

complement(2834..3724)

/gene="spym18.2230"

complement(2834..3724)

/gene="spym18.2230"

/note="best blastp match gb|AAC61484.1| (AF082738) ABC  
transporter (ATP-binding protein) [Streptococcus  
pyogenes]"

gene

CDS

/codon\_start=1  
/transl\_table=11  
/product="ABC transporter (ATP-binding protein)"  
/protein\_id="AAL98662.1"  
/db\_xref="GI:19749286"  
/translation="MIILKEAFLRLRNMSAIIEKKVTYNYHKOQEKPLDGVGS  
FHVKGWLSIIGHNGSGKSTIRLDIGLEPESGIIVDGLLTITNWEIRHKIGM  
VQNPDDQFVGAVVEDDVAFLENGKGAHEDIKERYNALELGMQNFKEPEPARLGS  
GQRVAVAGVAKFKIILIDEATSMIDPKRLEIKTIKNIDDIQVTVISITHDL  
DEVALSRLVMDKQGVSTSTPEQLFARGDELLOGLDIEFTTSVVVQMLQEGSYD  
YGYLKELENQLCLISKM"  
complement(3675..4217)  
/gene="spym18.2231"  
complement(3675..4217)  
/gene="spym18.2231"  
/note="best blastp match gb|AAK34824.1| (AE006636)  
phosphatidylglycerophosphate synthase [Streptococcus  
pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="phosphatidylglycerophosphate synthase"  
/protein\_id="AAL98663.1"  
/db\_xref="GI:19749287"  
/translation="MIKKENIPNLLTLVRIAMIPFLEFITSNNKVGWHIFAIVFAI  
ASFTDYLGDYLRKRWVSNFNGKGFADPLADKMLVMSAFIMLVGLVPAWVSAVLCR  
ELAVTGLRLLLVETGGKVLAAAMLKIKTKATQMLSTILLCHWIFLGNLLCIALFTT  
IYSGDYVEFKASFLKDFEK"  
complement(4231..5256)  
/gene="spym18.2232"  
complement(4231..5256)  
/gene="spym18.2232"  
/note="best blastp match gb|AAC61482.1| (AF082738) unknown  
[Streptococcus pyogenes]"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL98664.1"  
/db\_xref="GI:19749288"  
/translation="MSEKSLGEVRESRVGNKITLDDIELATGISSHYLLAMELDQFK  
IIPKEDQSLKEYADIVELDFNLKRRYQVNSKSDSKSVTEIVEEKLKRRQ  
EKLSHQSGDGGQPLEVSKNKTLLVPIKTVGLDLSKLPKTIKSPVQSKDNRLLPKK  
NYSRKNRNSDTSKSTLFPAILGTAFATVILVIGWQVQFQSKRAEVALPK  
ASEQSTGLTKNNSDRKTQVTEGSGNYLIATVKETVDIVSVLSEQAQSSWISLNN  
SEIGGGTITLQDTPYVATATPAETBSLTLTGTVNGSVNTIDGQPVLDLSPLTSTDLS  
YTFKIO"  
complement(5306..6595)  
/gene="spym18.2233"  
complement(5306..6595)  
/gene="spym18.2233"  
/note="best blastp match gb|AAC61481.1| (AF082738) unknown  
[Streptococcus pyogenes]"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL98665.1"  
/db\_xref="GI:19749289"  
/translation="MTKLKVINPNIDEDLYVVKLENGLTYYFIKKIGIEKTAFLVY  
GFGSDKLIVDSDRDPAGTAHFLEKLFDESGGDISLKFTQGAETNAFTTFNG  
TSFFTSASKFQNLQYFVLSANITDESYSREKKLIQOEDMDYDDADYRAYSGI  
LQNLPTKLANDIAGRESQTKTKILLTHTHTFYQPTNMSLFVIGDIDDEFLLA  
IQRTPTLSLPDRKRVTVDPHLYPIKSSSDVMDVTTAKLVGFRGIVLTLLQHSLLT  
YRIALKUFLKMGVTSKTYHTLYEDGIDDSFDVDVIEHNFQFVLISLDTPEFIAM  
SNYIKQFTKIKSKEFTNEHLNLLKEMYGDFIQSLDSIEHLTHOFSLYLSDSDKET  
YDPIPKIERTLTKDVVTIGKAFTEKADASDFTVFPK"  
complement(6597..7841)  
/gene="spym18.2234"  
complement(6597..7841)  
/gene="spym18.2234"  
/note="best blastp match gb|AAC61480.1| (AF082738) unknown  
[Streptococcus pyogenes]"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"

gene

CDS

Query Match	48.2%	Score 604.6;	DB 1;	Length 10851;
Best Local Similarity	68.7%;	Ref. No. 4.7e-143;		
Matches 848;	Conservative 0;	Mismatches 383;	Indels 3;	Gaps 1;
QY	10	TTAAAAAACCATCACTGTTGTGGCCTTAGTATTTTGGGTACTGTGATTACGTC	69	
Db	8237	TTTAAAAAACTTAAATGTGTTATCCCTTATTTTGGATATCTATCTTGATTATCTA	8296	
QY	70	AATGTTTATCTCTTTGGTGTCAAGAGACGTGTCAATTTATGGCTTTTGGTGATAGCT	129	
Db	8297	AATATGTATCTAATTTGGAC---ATCACTGTAGGAATTTATGGAGTAATTAATAACC	8353	
QY	130	TACCTATTAGTCAAAATGCTCTATCCCTTTTTTACAAGCCATTTAAGGAGAGCGCTGGG	189	
Db	8354	TATCTAGTTATCAAACTTGGATATCTTCTTTATGAGCCATTTAAGGAAGGCCAAT	8413	
QY	190	CAATATAGGTTGACGCCATTTATCCCTCTTTATACAAGATGCTGAGTCATTGCTAGAG	249	
Db	8414	GACTATAAGTGTGCTGTGAATTCCTTTATATGAAGATGCCGAGTCATTATTAGAA	8473	
QY	250	ACCTTAAAAAGTGTGACGAGCAAAACCTATCCCTAGCAGAAATTTATGTGTGACAT	309	
Db	8474	ACACTTAAAAAGTGTGTAGCACAGACCTATCCGTTATCAGAAATTTATTTGTTGATGAT	8533	
QY	310	GGAAGTGCTGATGAGACAGGTATTAGGCGCATTCAGAGACTATGTCGCTGACACTGGTGAC	369	
Db	8534	GGGAGTTCAAAACAGATGCAATACAAATTAATGAGAGTATGTAAATAGAGAAGTGGAT	8593	
QY	370	CTATCAAGCAATGTCAITTTGTCATCGGTGACAGAAAAATCAAGGAAAGCGCTATGCACAG	429	
Db	8594	ATTGTGGAAGCTTATGTCGTACCGTTCCTTGTCAATAAGGAAACGCCATGCTCA	8653	
QY	430	GCCTGGGCGCTTTGAAGATCAGACGCTGATGCTTTTTGACCGTTGACTCAGATACTTAT	489	
Db	8654	GGGTGGGCATTTGAAGATCTGACGTCGACGCTTTTTTTAAACGTAGATTACAGATACTTAT	8713	
QY	490	ATCTACCGTGAAGCTTTAGAGGAGTGTGTTAAAAACCTTTAATGACCCCACTGTTTTGCT	549	

RESULT 10	AE006637	LOCUS	9844 bp	DNA	linear	BCT 01-JUN-2001
DEFINITION	Streptococcus pyogenes M1 GAS strain SF370, section 166 of 167 of the complete genome.					
ACCESSION	AE006637 AE004092					
VERSION	AE006637.1 GI:13623172					
KEYWORDS	.					
SOURCE	Streptococcus pyogenes M1 GAS					
ORGANISM	Streptococcus pyogenes M1 GAS					
REFERENCE	Bacteria: Firmicutes, Lactobacillales; Streptococcaceae;					
AUTHORS	1 (bases 1 to 9844)					
TITLE	Perrettì,J.J., Mshan,W.M., Adjiç,D., Savic,D., Savic,G., Lyon,K.,					
	Ferreiro,C., Sezate,S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,					
	Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., Lin,S.,					
	Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.					
	Complete genome sequence of an M1 strain of Streptococcus pyogenes					
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)					
MEDLINE	21192684					
PUBMED	11296296					
REFERENCE	2 (bases 1 to 9844)					

## AUTHORS

Ferretti, J.J., McShan, W.M., Adijic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.

## Direct Submission

## JOURNAL

Submitted (10-APR-2001) Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, 940 SL Young Blvd, Oklahoma City, OK 73104, USA

## FEATURES

## Location/Qualifiers

1..3944

/organism="Streptococcus pyogenes M1 GAS"

/mol\_type="genomic DNA"

/strain="SF370"

/serotype="M1"

/db\_xref="taxon:160490"

246..1505

/gene="hasA"

/note="synonym: SPY2200"

246..1505

/gene="hasA"

/note="Best Blastp hit = pir|A53100 hyaluronate synthase A (HasA) - Streptococcus sp. (group A)"

>gi|473747|gb|AA17725.1| (L21187) hyaluronate synthase

[Streptococcus pyogenes]"

/codon\_start=1

/transl\_table=11

/product="hyaluronate synthase"

/protein\_id="AAK34828.1"

/db\_xref="GI:13623173"

/translation="MPIFKATLIVLSPFLISLILVNMVLESTGVIGVILITVLVIGLISFLYEPFKPHDKVAIPSYNEDAESLLETKSLVLAQYPSIETIYVDDGSNTDQILIEYVNVREIDICNVIVHRSVKNRKHQAQWAFERSADYPIVDSDDTYTPNALIELLKSFNDYVIAATGLNARNQTNLTITDIDYDNAGVRAAOSLTGNIYSGPLSIVREVIVPLNRYKNOFTGLPLPSIGSDRCLNVAIDIGRTVQOSTARCDDPFOKLSYLAQNRKNSFEKESIIISVKILSNPIVALTWIEVWPMMLIVAGNLLFNQALQDLKUPAFSLIFIVLCRNVHMKHPASLLSPGLGILHLFVLPQKLSYLSIKNTKMTWETGRKKVTIFK"

1541..2749

/gene="hasB"

/note="synonym: SPY2201"

1541..2749

/gene="hasB"

/EC\_number="1.1.1.22"

/note="Best Blastp hit = sp|Q07172|UDG\_STRPY UDP-GLUCOSE

6-DEHYDROGENASE (UDP-GLC DEHYDROGENASE) (UDP-GLCDH)

(UDPGDH) >gi|1075776|pir|A46089 UDPglucose

6-dehydrogenase (EC 1.1.1.22) - Streptococcus pyogenes

>gi|8569328|pdb|1DL1|A Chain A, The First Structure Of

UDP-Glucose Dehydrogenase (Udpghd) Reveals The Catalytic

Residues Necessary For The Two-Fold Oxidation

>gi|295189|gb|AA26899.1| (L08444) UDP-glucose hydrogenase

[Streptococcus pyogenes]"

/codon\_start=1

/transl\_table=11

/product="UDP-glucose 6-dehydrogenase"

/protein\_id="AAK34829.1"

/db\_xref="GI:13623174"

/translation="MKIAGSGYVGLSGLVLSLQNEVTIVDLPKSVKDKINNGSLP IQEYEVYKLSKLSIATKDSKAEALVIAATPYNRNLYNETDQVETVIVK EVLSVNSHATLILKSTIPICFITEMQKFDRIIFESPELSKALVDNLYPSRLIIV SCENSDPKVADAEKFAILLISAKKNVPVLMGASAEAVKLFANTYALKVAYF QDTIAEKLNSHMLIQGISDYDRGHYNNPSPFGYGLCPKDKTKQLYANNIP QTLTEAVSNRNKYSIAQIINVLREESFPVKVGYVRLMKNSDNFESAIDKV IDLAKSDIKIILYEPMLNKLESDQSVLVNDLENFKQANIIVTNRYDNLQDVNKK VYSRDFNRD"

2931..3845

/gene="hasC"

/note="synonym: SPY2202"

2931..3845

/gene="hasC"

/EC\_number="2.7.7.9"

/note="Best Blastp hit = sp|O54713|HASC\_STRPY

UTP-GLUCOSE-1-PHOSPHATE URIDYLTRANSFERASE (UDP-GLUCOSE

PYROPHOSPHORYLASE) (UDPGP) (ALPHA-D-GLUCOSYL-1-PHOSPHATE URIDYLTRANSFERASE) (URIDINE DIPHOSPHOGLUCOSE PYROPHOSPHORYLASE) >gi|995767|gb|AAA91810.1| (U03452) UDP-glucose pyrophosphorylase [Streptococcus pyogenes]"

/codon\_start=1

/transl\_table=11

/product="UDP-glucose pyrophosphorylase"

/protein\_id="AAK34830.1"

/db\_xref="GI:13623175"

/translation="MTKVKKAIIPAGLGTRELPAKAKEMLPIDVKTPIQFIVE ALKSGIEILVWIGKAKRSIEDHFNSELEYLNQAKNELKLVDETAINLHFIR QSHRGLGDAVLAQAKAFVGNPFVVMGLDGLDITNASAKPLTKQIMEDYDKTHAIR AVMKVPHEDEVSSYVIAQGVKAVGLYSDVTEFVEKPEQDAPSDLAIGRIYLLPETF GILRQTPGAGNEVOLTDADITLANKTORVFAREFKGNRYDVGKFGFKMTSIDVALSH PQVKEDLNKYLKLGKALEKSKVPTHSK"

4153..4566

/gene="SPY2203"

4153..4566

/gene="SPY2203"

/note="Best Blastp hit = pir|JC4754 hypothetical 13.6k

protein - Lactococcus lactis >gi|1107709|emb|CAA61548.1|

(X89367) orf121 [Lactococcus lactis]"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAK34831.1"

/db\_xref="GI:13623176"

/translation="MIYKLFTEFTTQALLKELGIQSGAIGKFLAETTVLFNGEDE RRRKRIIVGDKLSIPDDLIITIVPSQEEKEQAEEMAEKTRVAALYKQMNQANKK TSSKHNRRQSTTKSLRATKTKKTPAVRPPGI"

4568..5674

/gene="recF"

/note="synonym: SPY2204"

4568..5674

/gene="recF"

/note="Best Blastp hit = sp|P49999|REFC\_STRPY RECF PROTEIN >gi|1075773|pir|JC4077 recf protein - Streptococcus pyogenes >gi|533080|gb|AA85783.1| (U07342) Recf protein [Streptococcus pyogenes]"

/codon\_start=1

/transl\_table=11

/product="Recf protein"

/protein\_id="AAK34832.1"

/db\_xref="GI:13623177"

/translation="MMIKELKHRYNDHLLASFSSGLNVFIGNAOKTNFLAIV FLSTTRHRTADKELIHFDBSTVSLTGKIQISGTVLEINLSDKGRVTKNALQA KLSYICETMNVLEFAPEDQLYKCAPSLRKEKIDIDLGKIPYLSLSHNVHLKOR NSYIAKSAQIDRAFLAVLDEQLASIGARVNEHIDF INALEKANTHHQAISNGLESU SLUSYGSVVFDDKNTIYQOFLQLEKHNQKDFRKNTSGPHRDELAFYINGMANFA SOGQRSLIILSKMAEVSIMKALATGDNPIILLDDVMSLNDTRQTKLETIVKENVQT FTITTSLDLSQLEPIGHFWTKGTQVQDSDIH"

complement(5729..6562)

/gene="SPY2205"

complement(5729..6562)

/note="SPY2205"

/note="Best Blastp hit = sp|P40419|YCXE\_BACME HYPOTHETICAL 30.5 KD PROTEIN IN GDHI 5'REGION (ORF 2) >gi|801221|pir|J03084 hypothetical 30.5K chain - Bacillus megaterium >gi|216267|dbj|BAAL4098.1| (D90043) ORF2

[Bacillus megaterium] >gi|228656|pirf|1808263B gdh1

upstream ORF2 [Bacillus megaterium]"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAK34833.1"

/db\_xref="GI:13623178"

/translation="MFTWGSIGFVSNNIGKPKSQOTLGMTFGALLFLSLAVMLIVRPEM TLQLWFLGILGFGIMSIGQTFHMQYMGVSVANPLSSQSLVGLSIGLVFVHEWT RPFQVYVGSIALLLIIVGYFFSKQDDANAQVNLHNEKGRALTYSTIGVYAVL FNNMKFEVLSVILPMAYGMVGLAITFMSFKISIDQVYIKNSVVGGLMGIGNIFMLLA ASKAGLAIAFSFQSLGAISIVGGILFGLGETKTKEMRWVYTGIICTIVGAILLGTVK S"

complement(6794..8275)

gene

/gene="guab"  
 /note="synonym: SPY2206"  
 complement(6794..8275)  
 /gene="guab"  
 /EC\_number="1.1.1.205"  
 /note="The N-terminal amino acid sequence of this ORF has been determined from a spot isolated by 2-D gel electrophoresis from another strain of S. pyogenes. Hogan, D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen. 2000. Two-dimensional gel electrophoresis map of Streptococcus pyogenes proteins. Unpublished data.; Best Blast hit - pdb1zrj1A Chain A, Inosine Monophosphate dehydrogenase (Impdh; Ec 1.1.1.205) From Streptococcus Pyogenes"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Inosine monophosphate dehydrogenase"  
 /protein\_id="AAK34834.1"  
 /db\_xref="GI:13623179"  
 /translation="MSNWDKFLKGYTFDDVLLIPAESHVLPNEVDLTKLADNLT  
 NIPITAMDTVGTSMKAIARAGGLVHKNMSITEQAEVRKVRSENGVIDPE  
 FLPEHKYSEAEIMQRISSVPVETLANRKLVTITNDRMREISDYNAPISERT  
 SEHLVTAAGVGLDLETHRIEKLPLVDNSGRISGLTIKIDIEKVEFFRAKD  
 EFGRLVAAAGVTSDFEAEALPEAGADIVLDTHGHSAGVLRKIEHRAEPNR  
 TLGNIAIAGARALYDAGVDYKVGIPGSGICTIRYVAGVGPVQVTAIYDAAAVR  
 EYKTIADGKIKYSGDIVKALAGNVMGLGSMFAGTDEAPGETEYQGRKFKYRG  
 MGSIAKMGSDRIFGVSNEANKLVPEGLEGVKAYKGAASDIVFQMLGIRSGMY  
 VGADIDQLHENAOFVEMSGAGLIESHPHVDQITNEAPNYSVH"  
 complement(8583..9605)  
 /gene="trsa"  
 /note="synonym: SPY2207"  
 complement(8583..9605)  
 /gene="trsa"  
 /EC\_number="6.1.1.2"  
 Query Match  
 Best Local Similarity 48.0%; Score 601.4; DB 1; Length 9844;  
 Matches 846; Conservative 0; Mismatches 386; Indels 3; Gaps 1;  
 QY 10 TTAATAACCTATACCTGTTGGGCTTTAGTATTTTGGGTACTGTGATTACGTC 69  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 255 TTTAAAAAACTTTAATGTTTATCTTTTATCTTTTATCTTATCTGATTTATCTA 314  
 QY 70 AATGTTTATCTGTTGGTAAAGGAGCTTGTCAATTTATGCTTTTGTCTATGACT 129  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 315 AATATGTTATTTGGAC---ATCACTGTAGGATTTATGGAGTAATATTAATACC 371  
 QY 130 TACCTATTAGTCAAAATGCTTATCTCTTTTACAGCCATTAAAGGAGGCTGGG 189  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 372 TATCTAGTTATTAACCTGGATATCTTCTTTATGAGCCATTAAAGGAAAGCCACAT 431  
 QY 190 CAATATAAGCTGCAGCCATTATCTCTCTTATACAGAGATGCTGATGCTATGAG 249  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 432 GACTATAAGTTGCTGCTGATTAATCTCTTATTAAGAGATCCGAGTATTATTAGAA 491  
 QY 250 ACCTTAAAGTGTTCAGCAGCAAGCTATCCCTCTATACAGCAATTTATGTTGACGAT 309  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 492 ACTCTTAAAGTGTGTAGCAGACCTATCCGTTATCAGAAATTTATTTGTTGATGAT 551  
 QY 310 GGAAGTGTGATGACAGAGGTATTAAGCCATTGAGACTATGCTGTGACACTGGTAC 369  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 552 GGGAGTTCAACACACAGATGCAATTAATTAAGAGTATGTAATAGAGAGTGGAT 611  
 QY 370 CTATCAGCAATGCTATGTTCTATCGGTGAGAGAAATCAAGGAAGGCTATGACAG 429  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 612 ATTGTGCAAGCTTATGTTACCGTTCCTTGTCAATAAGGAAAGCCGATGCTCAA 671  
 QY 430 GCGTGGCCCTTTGAAGATCAGACGCTGATGCTTTTGTGACCGTTGACTCAGATCTAT 489  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 672 GCGTGGGCACTTTGAAGATCTGACGCTGACGTTTTTTTACCGTAGATCAGATCTAT 731  
 QY 490 ATCTACCGCTGATGCTTTAGAGGAGTTGTTTAAAAACCTTTAATGACCAACTGTTTTC 549  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 732 ATCTATCCAAATGCGCTTAGAAGAACTCTCTAAAAGCTTCAATGATGACAGAGTTTATGCT 791  
 QY 550 GCACGGGTCACCTTAATGTGAGAATAAGACAAACCAATCTCTTAACACGCTTGACAGAT 609  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 792 GCAACGAGGACATTTGAATGCTAGAACAGACAAACTAATCTATTAAACGACGATACAGAT 851  
 QY 610 ATTGCGTATGATAATGCTTTGGCTTGAACGAGCTGCCCAATCGGTTACAGATATATC 669  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 852 ATCCGTTAGATATGCTTTGGGTGGAGCGTCTGCTCAATCATTAACAGATATAT 911  
 QY 670 CTGTTTGTCTCAGTCCGCTTTACAGACGGGAGGTGTTGTTCTTCAACATAGAT 729  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 912 TTAGTTTGTCTCAGGACCAATTGATTTATCGAGTGAAGTGAATTTCTTACTACITAGAG 971  
 QY 730 AGATACATCAACAGACCTTCTGGGTATTCCTGTGAAGTATTTGGTGAAGACAGGTGCTG 789  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 972 CGTATAAAATCAACATCTCTAGTTTACCTGTTAGCAITGGGATGATCATGTTTA 1031  
 QY 790 ACCAATATGCAACTGATTTAGGAAGACTGTTTATCAATCCACTGCTAAATGATATACA 849  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1032 ACAATATGCTATGTTTATGAGCGACTGCTACCAATCAACAGCTAGATGATAT 1091  
 QY 850 GATGTTCTGCAAGATGCTTACTTACTTGAACGACAAACCGCTGGAACAGTCTTTC 909  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1092 GATGACCTTTTCCAATTTAAAAAGTTTATTAAGCAACAAATCGATGAATAAATCTTT 1151  
 QY 910 TTTAGAGAGTCCATTATTTCTTGAAGAAATCATGAACATCTTTTGTAGCCCTATGG 969  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1152 TTTAAGATCTATTTCTGTTAAAAAATTTCTTAAATCCCATCGTGCCTTATGG 1211  
 QY 970 ACCATCTAGTGGTCTGTTTATGATGCTTTTATTTCTGTGGTGGATTTCTTTGTA 1029  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1212 ACTATTTTGAAGTCTGTTATGTTATGATGTTGATTTGCGCAATTTGGATCTTTT 1271  
 QY 1030 GGCATGTCAGAGATTTGATTTGGCTCAGGCTTTAGCCCTTTCTGTTGATATCTTCAT 1089  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1272 AATCAGCTATTAATAGACCTTATTAACCTTTTGGCTTTTATCCATCATCTTATC 1331  
 QY 1090 GTTGCCTCTGTCGGAACATTCATTACATGTTAAGCAACCCGCTGCTCTCTTTGTTATCT 1149  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1332 GTTCTTTATGCTGATGTTCTTATGATGATGATGATGATGATGATGATGATGATGAT 1391  
 QY 1150 CGTTTATGAGGTTGCTGATTTGTTGCTCCTACAGCCCTTGAATATATATCTCTTTT 1209  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1392 CCTCTGTATGAATATATACATGTTGTTGCTTACAGCCCTTAAACCTTTATCTTTATGC 1451  
 QY 1210 ACTATTAGAAATGCTGACTGGGACACGTAATAA 1244  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 1452 ACCATTAATAATAGGAATGGGACACGTAATAA 1486  
 RESULT 11  
 AE014172  
 LOCUS  
 DEFINITION  
 Streptococcus pyogenes MGAS315, section 37 of the complete genome.  
 ACCESSION  
 AE014172 AE014074  
 VERSION  
 AE014172.1 GI:21905566  
 KEYWORDS  
 Streptococcus pyogenes MGAS315  
 Streptococcus pyogenes MGAS315  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 ORGANISM  
 1 (bases 1 to 53896)  
 Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S.,  
 Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F.,  
 Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schliever, P.M. and  
 Musser, J.M.  
 TITLE  
 Genome sequence of a serotype M3 strain of group A Streptococcus:  
 Phage-encoded toxins, the high-virulence phenotype, and clone  
 emergence  
 JOURNAL  
 Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)  
 REFERENCE  
 2 (bases 1 to 53896)

## AUTHORS

Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,  
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,  
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and  
Musser,J.M.

## Direct Submission

Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,  
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,  
Hamilton, MT 59840, USA

## FEATURES

## Location/Qualifiers

1..53896

/organism="Streptococcus pyogenes MGAS315"

/mol\_type="genomic DNA"

/strain="MGAS315"

/serotype="M3"

/db\_xref="taxon:198466"

/notes="group: A"

/complement(200..1480)

/gene="hiss"

/note="synonym: SpyM3\_1815"

/complement(200..1480)

/gene="hiss"

/note="best blastp hit: gb|AAL98633.1| (AE010120) putative  
histidine-tRNA ligase [Streptococcus pyogenes MGAS8232]"

/codon\_start=1

/transl\_table=11

/product="putative histidyl-tRNA synthetase"

/protein\_id="AAM80422.1"

/db\_xref="GI:21905567"

/translation="MKLQKPGTQDILPCDAKQWYVESVARDTSQNYGEITPMF  
EHEVLSRSGVTDTIVKEMDFDKGDRHITLRPEGTAPVRSIVENKULFAPEVOK  
PVKITYGSMFYERPGQRLREFHQVEGFGAANPATDVETIAMVHLEFKIGIND  
VTLHLSGFSRSAYRQALIDITPMRDLSQSRDLRDNPLRVDSKEKEDLIA  
VERAPILLDYDEESQAFKAMEALDIPYVIDTNMVRGLDYNHTIFEITISVE  
GSOLITCGDYDSLVGFGPGTGFGLERLLMVEIKOGITLPIETEMDVILA  
VLGDGANSKALEVQALRROGFTAEADYLRGKIRAKFSADTFKAKLVMTLGESEVEA  
GKAVIKNRSRQVEVSEFEDMTNPNANISEQLLS"

## gene

/gene="rpmf"

/note="synonym: SpyM3\_1816"

/1700..1882

/gene="rpmf"

/note="best blastp hit: gb|AAL98635.1| (AE010120) 50S  
ribosomal protein L32 [Streptococcus pyogenes MGAS8232]"

/codon\_start=1

/transl\_table=11

/product="50S ribosomal protein L32"

/protein\_id="AAM80423.1"

/db\_xref="GI:21905568"

/translation="MAYPARHTSKAKNKRTHYKLTAPSVQFDETTGYSRSHVSL  
KGYKGRKIAKANEA"

## gene

/gene="rpmg"

/note="synonym: SpyM3\_1817"

/1898..2047

/gene="rpmg"

/note="best blastp hit: gb|AAK34796.1| (AE006634) 50S  
ribosomal protein L33 [Streptococcus pyogenes M1 GAS], and  
gb|AAL98636.1| (AE010120) 50S ribosomal protein L33  
[Streptococcus pyogenes MGAS8232]"

/codon\_start=1

/transl\_table=11

/product="50S ribosomal protein L33"

/protein\_id="AAM80424.1"

/db\_xref="GI:21905569"

/translation="MRVNITLHESGERLYTSKKNRTPDRQLKKYSPKLRKHT  
FTEVK"

## gene

/gene="cadD"

/note="synonym: SpyM3\_1818"

/2328..2942

/gene="cadD"

/note="best blastp hit: gb|AAK34797.1| (AE006634) putative  
cadmium resistance protein [Streptococcus pyogenes M1

## GAS]"

/codon\_start=1

/transl\_table=11

/product="putative cadmium resistance protein"

/protein\_id="AAM80425.1"

/db\_xref="GI:21905570"

/translation="MIONVPSIILYSSTAVDLIILMLFFAKRKRODINIYLGOF  
LGSVSLILSLFAFVDYIPSKELGLGLIPFGLKVLGLDSDGSAIAKEGLSK  
DNKNLILVAMITFASCGADNIGVFYFTTLNLANLIVALLFLVMITLLVFSQAKL  
AQVPSVGETLEKYSRWFIAVVLGLMYILIIENNSFDMIAVLG"

2954..3292

/gene="cadX"

/note="synonym: SpyM3\_1819"

2954..3292

/gene="cadX"

/note="best blastp hit: gb|AAK34798.1| (AE006634) putative  
cadmium efflux system accessory protein [Streptococcus  
pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="putative cadmium efflux system accessory  
protein"

/protein\_id="AAM80426.1"

/db\_xref="GI:21905571"

/translation="MKDSICQGVINOQNVTTATNYLEKEKVKOKSLRILSKFTDNKO  
INIIFLLAYEELCVCDIACLLNLSMASASHLRLKANQNILDRREGKIIYFFIKDE  
EIRDFENQLG"

3343..4266

/gene="SpyM3\_1820"

3343..4266

/gene="SpyM3\_1820"

/note="best blastp hit: gb|AAK34799.1| (AE006634)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAM80427.1"

/db\_xref="GI:21905572"

/translation="MKFFENYSOEIPIRIKLNKKHNITOSELGNAGOVSOVESGKR  
PITSSMLVNLALTASSYIVIVGELDEFIENLHVFPSSILYVDELEAVDENLYSPMS  
DDLISQSSCLSIKTFANFIQKRPMISTETEMDTFHKDDIDVWVGESYNPARS  
FRNNPILNIVDFEEMADILLMLRNLIRSFENVCNTLFDKXNGEPTFNLDKI  
DSINKMWSNSVSTEIIPNLIKKLRNPENFENIGFVMDILERMFKEDIPKSYLTSVPL  
VSKKASTLSISNGSQKIDKSKFAQLSL"

4332..5084

/gene="SpyM3\_1821"

4332..5084

/gene="SpyM3\_1821"

/note="best non-GAS blastp hit: gb|AAK79835.1|AE007695\_8  
(AE007695) FtsK-like DNA segregation ATPase, YDCQ  
B. subtilis ortholog [Clostridium acetobutylicum]"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAM80428.1"

/db\_xref="GI:21905573"

/translation="NAIPDLKFSNLSQKIPKIKHFLNLDKGSWIAIGNSGSK  
PYATYFLSVLKMSGLIIDPKRDTSPQWAKENIAVHVHNHKSDFVSQVNEOL  
NOCATLIQKQOAILDYNPHOFTHLTIVIDEVALSEGVKNKEAFSLISQIALIG  
HATKHLFLQSQRFDNHTIPISVREQLNVLQIGINOKTQQLFDPDPEGIVIPG  
HGTGLIQVDNEHSYQVLEPLLCPTIYTKRGIL"

5081..5683

/gene="SpyM3\_1822"

5081..5683

/gene="SpyM3\_1822"

/note="best blastp hit: gb|AAK34801.1| (AE006634)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAM80429.1"

/db\_xref="GI:21905574"

/translation="MTTKTYPMRSFNLIYLLLIAPICVYFGWLFNLVSVQLSTVTFV

```

NLDVKNVNTDISSVCYGFALISLLALLSLKMIQRLNDSIGNYCOSIYRTFLIRH
FLFORERQVKNSEHOTVTTINPHNGFNRAVRKICVIDRKDAITVELKVPDQCGQ
KILKDTGVQLKEEIVIGHDYFESSPVRNRQLWFGQR"
complement(6159..6323)
/gene="SpyM3_1823"..6323)
complement(6159..6323)
/gene="SpyM3_1823"
/note="best blast hit: gb|AAK34556.1| (AE006610)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA080430.1"
/db_xref="GI:21905575"
/transl_table="MLFGTLLIAISYIPNSLWGWGYPVAILWLIIMRSALRIERE
KTYDQVQTP"
complement(6651..6809)
/gene="SpyM3_1824"
complement(6651..6809)
/gene="SpyM3_1824"
/note="best blast hit: emb|CAB65345.1| (Y18528) PR7
protein [Plasmodium falciparum]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA080431.1"
/db_xref="GI:21905576"
/transl_table="MFLNNTFNFEIPLKHVFIKSSLIYKVVNKKYKDKCPYDNY
QKDKRKA"
complement(6894..7736)
/gene="SpyM3_1825"
complement(6894..7736)
/gene="SpyM3_1825"
/note="best non-GAS blast hit: gb|AAK99693.1| (AE008462)

Query Match          47.7%; Score 598.2; DB 1; Length 53896;
Best Local Similarity 68.3%; Pred. No. 2.1e-141;
Matches 844; Conservative 0; Mismatches 388; Indels 3; Gaps 1;

Qy 10 TTAATAAACCTCATACTGTGGCTTTAGTATTTTGGGTACTGTGTATTACGTC 69
Db 33882 TTTAAATAAACCTTAAATGTTTATCTTTTATTTGTTGATATCTATCTATCTA 33941
Qy 70 AATGTTTATCTCTTGGTGTCAAGAGAGCTGTGCAATTAATGCTTTTGTCTATAGCT 129
Db 33942 AATATGTATCTATTTGG---AACATCACTGTAGGAATTTATGGAGTAATATTAATACC 33998
Qy 130 TACCTATTAGTCAAAATGCTCCTTATCTTTTACAGCCATTTAAGGAGAGGCTGG 189
Db 33999 TATCTAGTTATTAACCTGGATTAATCTTCCCTTATGAGCCATTTAAGGAATCCACAT 34058
Qy 190 CAATTAAGTTGACGCAATTAATCCCTCTTATAACGAAGATGCTGAGTCATTGCTAGAG 249
Db 34059 GACTATTAAGTTGCTGTGAATCTCTTATAATGAAGATCCGAGTCATTATTAGAA 34118
Qy 250 ACCTTAAGAGTTTCACAGCAAACTATCCCTAGCAGAAATTTATGTTGTTGACAGAT 309
Db 34119 ACCTTAAGAGTTGTTAGCAGACACTATCCGTTATCAGAAATTTATATTTGTTGATGAT 34178
Qy 310 GGAAGTCTGATGACAGAGTATTAAAGCGCAATTGAAGACTATGTGGTGACACTGGTAC 369
Db 34179 GGGAGTTCAACACAGATGCAATACAAATTAATGAAGATGATTAATAGAGAGTGGAT 34238
Qy 370 CTATCAAGCAATGCTATGTTTCATCGGTGACAGAAATCAAGGAAGCGTCATGCACAG 429
Db 34239 ATTTGTGCAAGCGTTATGCTTCACCGTTCCCTTGTCAATAAAGAAAGCGCAATGCTCAA 34298
Qy 430 CCCTGGCCCTTTCAAGATACAGCGTGTATGCTTTTGGACCGTTGACTACAGATCTAT 489
Db 34299 GCGGGGCACTTTGAAGAATCTGACGCTGACGTTTTTTTTTAAACCGTAGATCTAGATCTAT 34358
Qy 490 ATCTACCCCTGATGCTTTTAGAGAGTGTGTTAAACCTTTAATGACCAACACTGTTTGGCT 549

```

```

Db 34359 ATCTATCCAAATGCCTTAGAAGAACTCTCTAAAGCTTCAATGATGAGACAGATTTATGCT 34418
Qy 550 GCGAGGGTCACTTAAATGTCAGAAATAGACAAACCAATCTCTTTAAACACGCTTGACAGAT 609
Db 34419 GCAACAGGACATTTGAATGCTAGAAACAGACAAACTAATCTATTAAAGCGACTTACAGAT 34478
Qy 610 ATTCCGTATGATAATGCTTTTGGCGTTGACAGAGCTGCCCAATCCGTTACAGGTAATATC 669
Db 34479 ATCCGTTAGATAATGCTTTTGGGGTGGAGCGTGTCTCAATCATTAACAGGTAATAT 34538
Qy 670 CTGTTTGTCTCAGGTCCGCTTAGCGTTTACAGACGCGAGGTGGTTTCTTCAACATAGAT 729
Db 34539 TTAGTTTGTCTCAGGACCATTTAGTATTATTCGACGCTGAAGTGAATATTCCTACACTAGAG 34598
Qy 730 AGATACATCAACAGACCTTCTGGGTATTCCTGTGAAGTATTGGTGAATGACAGGTCGTTG 789
Db 34599 CGCTATAAANAATCAACATCTCTAGTTTACCTGTGTAGCAITGGGGATGATGATGTTA 34658
Qy 790 ACCAACTATGCAACTGATTAGGAAAGACTGTTTATCAATCCACTGCTAAATGTAATACA 849
Db 34659 ACAAAATATGCTATGATTAGGACGCACTGTCTACCAATCAACAGCTAGATGATGAT 34718
Qy 850 GATGTTCTGACAAAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAACAAGTCCTTC 909
Db 34719 GATGACCTTTTCCAAATTAATAAGTATTATTAAGCAACAAATTCGATGGAATAATCTTTT 34778
Qy 910 TTTAGAGATGCTATTTCTTCTTAAGAAATCATCAACAATCTTTTGTAGCCCTATGG 969
Db 34779 TTTAGAGATGCTATTTCTCTCAAAAATCTTTCTTAATCCCATCGTGGCTTATGG 34838
Qy 970 ACCAATCTGAGGTGCTATGTTTATGATGCTGTTTATTTCTGTGGTGAATTTCTTTGTA 1029
Db 34839 ACTATTTTTCGAAGTCGTATGTTTATGATGCTGATTTGCGCAATTTGGAATCTTTGTTT 34898
Qy 1030 GGCATGTCAGAGAAATTTGATTTGGCTCAGGGTTTATAGCCTTTCTGGTGATATCTCAAT 1089
Db 34899 AATCAAGCTATTCAATTAGACCTTATTAACCTTTTGGCTTTTATCCATCATCTTATC 34958
Qy 1090 GTTGCCCTGTCGGAACATTCATTACATGCTTAAGCAGCCGCTGCTCTCTGTTATCT 1149
Db 34959 GTTGCTTTATGCTGTAATGCTTATTAATGCTCAAACTCTGCTAGTTTGTGTTATCT 35018
Qy 1150 CGCTTTTATGGGTGCTGCAATTTGTTTGTCTACAGCCCTTGAATATATATCTCTTTT 1209
Db 35019 CTTCTGTATGGAATATTACACTTTGTTGCTTACAGCCCTAAACCTTTATCTTTATGC 35078
Qy 1210 ACTATTAGAAATGCTGACTGGGGAACAGTAAAAA 1244
Db 35079 ACCATTAAAAATAGGAATGGGGAACAGTAAAAA 35113

RESULT 12
AP005146
LOCUS 323825 bp DNA linear BCT 13-JUN-2003
DEFINITION Streptococcus pyogenes SSI-1 DNA, complete genome, section 6/6.
ACCESSION AP005146 BA000034
VERSION AP005146.1 GI:28811737
KEYWORDS
SOURCE Streptococcus pyogenes SSI-1
ORGANISM Streptococcus pyogenes SSI-1
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1
AUTHORS Nakagawa,I., Kurokawa,K., Yamashita,A., Nakata,M., Tomiyasu,Y.,
Okahashi,N., Kawabata,S., Yamazaki,K., Shiba,T., Yasunaga,T.,
Hayashi,H., Hattori,M. and Hamada,S.
TITLE Genome Sequence of an M3 Strain of Streptococcus pyogenes Reveals a
Large-Scale Genomic Rearrangement in Invasive Strains and New
Insights into Phage Evolution
JOURNAL Genome Res. 13 (6), 1042-1055 (2003)
MEDLINE 22683278
AUTHORS 2 (bases 1 to 323825)
Yamashita,A., Nakagawa,I., Kurokawa,K., Nakata,M., Tomiyasu,Y.,

```



Query Match	47.7%;	Score	598.2;	DB 1;	Length	323825;
Best Local Similarity	68.3%;	Pred. No.	2.2e-141;			
Matches	844;	Conservative	0;	Mismatches	388;	Indels 3; Gaps 1;
QY	10	TTAAAAACCTCATACACGTGCTGGCCCTTTATGTTATTTTGGTACTGTTGATTGATTCAGTC	69			
Db	303842	TTTAAAAACCTTTAATGTGTTTATCCCTTTATTCCTTTATTTGTTGATACTCATCTTGATTTCCTA	303901			
QY	70	AATGTTTATCTCTTTGGTGCTTAAGAGACCTTGCTCAATTTATGGCTTTTGTCTGATAGCT	129			
Db	303902	AATATGTATCTATTGCG---AACATCAACTGTAGGAATTTATGAGTAATAATTAAATAACC	303958			
QY	130	TACCTATTATGCTCAAAATGTCCTTATTCCTTTTTCACAGCCATTTAAGGAGAGCGCTGGG	189			
Db	303959	TATCTAGTTTATTAACCTTGGATTATCTTCCTTTATGAGCCATTTAAGAGAAATCCACAT	304018			
QY	190	CAATATAAGTTGCGACCATTTATTCCTCTTTATACGAAGATGCTGAGTCATGCTAGAG	249			
Db	304019	GACTATAAGTCTGCTGCTGTAATTCCTCTTTATAATGAAGATGCCGAGTCATTATTAGAA	304078			
QY	250	ACCTTAAAAAGTGTTCAGCAGCAAAACCTATCCCTAGCAGAAATTTATGCTGTGTGACGAT	309			
Db	304079	ACACTTAAAGTGTGTAGCACAGACCTATCCGTTATCAGAAATTTATTTGTTGTATGAT	304138			
QY	310	GGAGTGTGATGAGACAGGTATTAGCGCATTAAGACTATGTGGCTGCACACTGGTGAC	369			
Db	304139	GGAGTGTCAACACAGATGCAATACAAATTAATGAAGATGTAAATAGAGAGTGGAT	304198			
QY	370	CTATCAGCAATGTCATGTTTCATCGCTCAGAGAAATCAAGAAAGCGCTGTGCACAG	429			
Db	304199	ATTGTGCAACGTTATGCTTTCACCGTTCCTTGCTCAATAAGAAAGAACGCCATGCTCAA	304258			
QY	430	GCTGGGCGCTTTGAAGATCAGACGCTGTGCTTTTTTGACCGTTGACTCAGATACTTAT	489			
Db	304259	CGGTGGGCAATTTGAAGATCTGACGCTGACGTTTTTTTACCGTGAATTCAGATACTTAT	304318			
QY	490	ATCTACCGCTGATGCTTTAGAGAGTGTGTTAAAAACCTTTATAGACCAACTGTTTTGCT	549			

FEATURES	Location/Qualifiers	Accession
source	1..1893	AF065404
gene	/organism="synthetic construct"	AF065404.1
	/mol_type="genomic DNA"	GI:4894216
	/db_xref="taxon:32630"	
gene	<1..>1893	
	/gene="has operon deletion mutant"	
gene	<1..>423	
	/gene="hasA"	
misc_feature	430..1137	
	/note="hyaluronate synthase"	
gene	/note="derived from vector pTW100"	
	<1138..1356	
	/gene="hasB"	
gene	/note="UDP-glucose dehydrogenase"	
	1406..>1893	
	/gene="hasC"	
CDS	1406..>1893	
	/gene="hasC"	
	/codon_start=1	
	/transl_table=11	
	/product="UDP-glucose pyrophosphorylase"	
	/protein_id="AAN59902.1"	
	/db_xref="GI:24415826"	
	/translation="MTKVKRAIIPAAAGLGRPLPAKALAKEMLPIDVKPTQIFVEE ALKAGIEILVVTGKARSIEDHFDNSFEYNLQAKSTELLKLVDETTAINLHFIR QSPRGLGDAVLAQAKAFVGNPFVWMLGDDLDITNFSAPLTKLIEDYDCTHASTI AV"	
BASE COUNT	559 a 427 c 385 g 522 t	
ORIGIN		
Query Match	32.6%; Score 409.2; DB 12; Length 1893;	
Best Local Similarity	92.9%; Pred. No. 2.7e-93;	
Matches	429; Conservative 0; Mismatches 33; Indels 0; Gaps 0;	
QY	475 GACTCAGACTATATATACCTGATGCTTTAGAGAGTGTGTTAAACCTTTAATGAC	534
Db		
QY	1 GACTCAGACTATATATACCTGATGCTTTAGAGAGTGTGTTAAAGACCTTTAATGAC	60
Db		
QY	535 CCAACTGTTTCTCGACGGGTCACCTTAATGTCAGAAATAGACAAACCAATCTCTTA	594
Db		
QY	61 CCAACTGTTTCTCGACGGGTCACCTTAATGTCAGAAATAGACAAACCAATCTCTTA	120
Db		
QY	595 ACACGCTTGACAGATATTCGCTATGATAATGCTTTGGCTTGAACGAGTGCACCAATCC	654
Db		
QY	121 ACACGCTTGACAGATATTCGCTATGATAATGCTTTGGCTTGAACGAGTGCACCAATCA	180
Db		
QY	655 GTTACAGTAATATCTGTTTGGCTACAGTCCGCTTACGAGCGAGTGTT	714
Db		
QY	181 GTTACAGTAATATCTGTTTGGCTACAGTCCGCTTACGAGCGAGTGTT	240
Db		
QY	715 GTTCTACATAGATAGATACATCAACGACCTTCTCGGTATTCCTGTAAGTATCGT	774
Db		
QY	241 GTTCTACATAGATAGATACATCAACGACCTTCTCGGTATTCCTGTAAGTATCGT	300
Db		
QY	775 GATGAGGTGCTTGACCACTATGCAACTGATTTAGGAAGACTGTTTATCAATCCACT	834
Db		
QY	301 GATGAGGTGCTTGACCACTATGCAACTGATTTAGGAAGACTGTTTATCAATCCACT	360
Db		
QY	835 GCTAAATGATATACAGATGTCCTGCAAGATCTCTACTTACGAGCAGCAAAACCGC	894
Db		
QY	361 GCTAAATGATATACAGATGTCCTGCAAGATGTCCTACTTACGAGCAGCAAAACCGC	420
Db		
QY	895 TGGAAAGTCTCTCTTTAGAGAGTCCATTTATTTCTGTTAAG	936
Db		
QY	421 TGGGATCCACTAGTCTTAGAGCGCGCCGCCCGGCTGGAG	462
Db		
RESULT 14		
AF065404	181654 bp DNA circular BCT 20-OCT-1999	
LOCUS		
DEFINITION	Bacillus anthracis virulence plasmid pX01, complete sequence.	

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL  
MEDLINE  
PUBMED

REFERENCE  
AUTHORS

TITLE  
JOURNAL

FEATURES  
Source

CDS

RBS

CDS

CDS

CDS

AF065404  
AF065404.1  
GI:4894216

Bacillus anthracis  
Bacillus anthracis  
Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus; Bacillus cereus group.

1 (bases 1 to 181654)  
Okinaka,R.T., Cloud,K., Hampton,O., Hoffmaster,A.R., Hill,K.K., Keim,P., Koehler,T.M., Lamke,G., Kumano,S., Mahillon,J., Manter,D., Martinez,Y., Riche,D., Svensson,R. and Jackson,P.J.  
Sequence and organization of pX01, the large Bacillus anthracis plasmid harboring the anthrax toxin genes  
J. Bacteriol. 181 (20), 6509-6515 (1999)

2 (bases 1 to 181654)  
Okinaka,R.T., Cloud,K., Hampton,O., Hoffmaster,A., Hill,K., Keim,P., Koehler,T., Kumano,S., Lamke,G., Manter,D., Martinez,Y., Riche,D.O., Svensson,R. and Jackson,P.  
Direct Submission  
Submitted (14-MAY-1998) Life Sciences Division, Los Alamos National Laboratory, TA43, IS-6, HRL-1, MS M888, Los Alamos, NM 87545, USA

1..181654  
/organism="Bacillus anthracis"  
/mol\_type="other DNA"  
/strain="Sterne"  
/db\_xref="taxon:1392"  
/plasmid="virulence plasmid pX01"  
/note="Hypothetical protein, Methanobacterium thermoautotrophicum, (AE000831)"  
/codon\_start=1  
/transl\_table=11  
/product="pX01-01"  
/protein\_id="AAD32305.1"  
/db\_xref="GI:4894217"  
/translation="MEVLEFELILAVLPLNSVKKVKPKGKAGEKLVKRLMSKL  
DKPSYLVHNVYTYEYDITQIDHIVIAETGVVETKNEYGWIYSEKAAHWTOGI  
FKKSFONPHONYKHAIKAEIWLTEQLPCISMAAFHPKGLKRVNVHSEKHKVYL  
HNKFAQ"  
complement(1639..1644)  
2751..2755  
2760..3338  
/codon\_start=1  
/transl\_table=11  
/product="pX01-02"  
/protein\_id="AAD32306.1"  
/db\_xref="GI:4894218"  
/translation="MLEHTILFSLVQLVPLIYVSTGIILILELTDENKWTSEYGLF  
AIFLLIFQVFNKASIIISLQHEKWFSLNIFLVSFLVLPFASIPTLTKTAFILQM  
TYLFLLEYLCLIDRLPHNLYRVKELICDKKQKIERIIPNNEEGHTTITKTRDEQ  
YFLYAVYLSGLISPISEVEFEYERNLMGR"  
complement(3389..4024)  
/codon\_start=1  
/transl\_table=11  
/product="pX01-03"  
/protein\_id="AAD32307.1"  
/db\_xref="GI:4894219"  
/translation="MKREKMGIVLCVLVLIILLVPLQAAQSVKVDENKLGNDYS  
SSFQGLNNRTKTEKDFSGSIIOKIEGLYAPAFKVGYNPISFVLGIVVMVAL  
ITKNGWNAWTKGSMIFSFISMSMRLSVYELFSSDVSFNPITFDLDFKATALLF  
TPFMLLAGIRLRHIESIKQPEYRWANRVITGSCLLCLAVLAPWLFANM"  
complement(4024..4653)  
/codon\_start=1  
/transl\_table=11  
/product="pX01-04"  
/protein\_id="AAD32308.1"  
/db\_xref="GI:4894220"  
/translation="MNINSISNNVKELEQTVYRKQVQVLRVSKSSNIHAYNIFP  
EFSEVLSNTELDSTSVKVIDRKLSQLQEKHNSFIMTCVEEDGVIDDE  
TIFNTYVVPPTIANKVKLVCGLGRLDLSKLHITVLNNINRIETEFYENENY"

```

RBS
CDS
VAYEKEIANATMGAYTKANIYYAADLYAAAGNEIKLAKSIKPKLAGDV
complement(4034. .4039)
complement(4660. .4665)
complement(4989. .5309)
/codon_start=1
/transl_table=11
/product="px01-05"
/protein_id="AAD32309.1"
/db_xref="GI:4894221"
/transl_table="MELLEDYMQOEVSIPETPEVFPVQOQVADVADKMDVFWRT
TKVLEQEQGLFVIMNSNNRYWHIEGEILLPENQLLYMKLEVDSPSYITVLEWSNRQ
LGN"
complement(5318. .5323)
complement(5720. .6220)
/codon_start=1
/transl_table=11
/product="px01-06"
/protein_id="AAD32310.1"
/db_xref="GI:4894222"
/transl_table="MLGDTGDLTLLIQNIKSQNLKGSTTEWTAKASGESKRVLANI
GNYITVFESIEIQTNNVEKLPGTMDVAFMKEAQFQKAVECIKKEQVPCVRKQNV
LEVDGEIEDLRALTTLNTDKMLVKQDITFDSVIVEVDPSSKKKGKKNKIIDQY
EFDLPC"
complement(6231. .6236)
complement(6544. .8352)
/note="reverse transcriptase, IS629, E. coli, (AB011549);
reverse transcriptase, IS629 (574 aa) E. coli (AB011549);
306/543 positive aa (56%)"
/codon_start=1
/transl_table=11
/product="px01-07"
/protein_id="AAD32311.1"
/db_xref="GI:4894223"
/transl_table="MNTVNMKSAVSPQVOTOWNSTOMKENTYVKKLORIVRAEQFGN
KRVKQLQMLRSGKALLSLIKRVTOINQCKRTAGIDGITTPEDVRKLEHLIAGY
SVNIKARPKRVKIPKNGKKRPLGIVIKDRFQNMVKNALPQWECRPFMSNGF
RPSRSHDAMANLFLKSRTNRNMFEGDQFCFQNLNHEILSCIEGPPYSNAINQ
WLAGCINDDFVYCTGPGGIIISPLLANIALHGEKELGVRVHPFKRGGAMLYPD
SLGIYAADFPVYCNSEFAESMTAKIQLPYDLRGLAKAEKTRVHTIDGDFDIFG
NFQYPTPEGSQLFKPSQNSQVKKAKKISEIPKSHRGRSIGOLIRKLNVPVIGIANY
WSPVAKVITYGIDSYVHKVHHLKHKHKGARWLNKKYPHDPTGVSDQKMLTD
PDNHKQLIRMTPIVRHVLIKYKNSPDDANLKKYFAKDEKIFNRFNNSKRLAK
KTYKRCVNSLSGEEPLESHNIPVPIVGGDKDNDLLELLHSCQHQHALLWEYGN
GKQLPKVQYLRKQKIPFNKSKAVRTMLNTPKFKY"
complement(8362. .8367)
complement(8956. .11541)
/codon_start=1
/transl_table=11
/product="px01-08"
/protein_id="AAD32312.1"
/db_xref="GI:4894224"
/transl_table="MLQEIHKQDTFDBQTPAILQLETGTASHSAFCFAMAINHREI
QFVAVGANDSLKSFRAAISMTGTSRYFEGCKRESHYTLDKVNVISKQOFFINTQ
TINKKAITATSKOLEPKYIVAIDEAPMOVDPFLMPPYGLPILFEWAKPIYEMLT
KNLLOPLEVDKNEFSSLSIAQVLEKDCKEFLESMIRSKCOFFQEGTGKINKI
ODLNEVLEYSPVMDVYKLEDEPLHQPKQZELSHFDIKRPLFPVQAHVATGAAGA
LQVQGLIIGEMSSGSIMTAVDGYFHLTGKGRVTCFVPPVPTLEKWAKEEIRH
LPDAVHLIRKTEDLIRHQSWIQTGRPKPEPTFVISTTHRGDSIKQMPLYKQ
IALSKRSVEQRYKNGYPCDCAKLRKTSIMVHQAGEQKVCQYKDFGSDL
DKNKNSVCADKNSWSPKVTYKASFKDWTYKYLQVAKGKPKQKOLEN
RVKPYDAGQSRVARKVAVETVIRKMKHPEDALICDEVHELKSETAQGTALGSLKV
CKVAVAGTFLGGASDVYILRFLPQKAVKSGFEYSSEMMNETYIGNIETVYHD
GEVTSNARSAGQGLDKTVVPGISPVTFQFMDTTINVELKDVMPNVELVNPYD
LVDMSQEQKAYEHMKSEFKAIEAAGTDPAGKMLSYINTGNAYPDNMIHPNYI
LDAGKEIWNDRKDFEDCQPKELQSLIRTEISGRPTIYVCDQTSVKRERO
VQPRLKQVEQIPGAKVATLRTNTVTPPKRSEMLKEQVNSGVNIIICSLVNVGLDC
ATSR"
complement(11549. .11554)
complement(11613. .11840)
/codon_start=1
/transl_table=11
/product="px01-09"
/protein_id="AAD32313.1"

```

```

/db_xref="GI:4894225"
/transl_table="MSVLLIGSLLIWFESNRMERIAVLTVICLASCALGAGTHLE
VLEQFSVPHGVGDMLSAAQSLPEWKRFDIF"
complement(11847. .11852)
complement(11856. .12947)
/codon_start=1
/transl_table=11
/product="px01-10"
/protein_id="AAD32314.1"
/db_xref="GI:4894226"
/transl_table="MRLGNDLAAQFYPTPLTEGKHLVRLQMBSERAYACDPQCGRS
TLISFADETKQAGVQCHTYVELDANRYGQAKOEOLOVYVRSFSSPMISHNYPILF
LNPPYNTLETENAKSEKMEFELKRAHLYLQEGSIWYVPTDFRARDODISYFLAKH
VEELGLMRFGEDENEEFQKOCYFVIGKRATKNDITYFNDRFANFENNDSDDDFVRKH
VSTLAQWNEKOWTPELRHONKLIPTSRVDYKDAYKGKSGEIASLAKKRLNNGY
LLQCKTAARAAKMPIASQGLGLLITGVADLLGEGDTLHVRGSENYVTEHSFEQL
EKTTEVYLQKRRAKFITATPTGVVVELV"
complement(12956. .12961)
complement(13209. .13562)
/codon_start=1
/transl_table=11
/product="px01-11"
/protein_id="AAD32315.1"
/db_xref="GI:4894227"
/transl_table="MEDNOKSYAAPVMSSSESRARVNMRCGEAEIILPDCKEYFEP
HALNEKATQETALXIEIQEMKYGIVILWRFVGOQSYSGTKIVAEKFKILRUK
DYFDYEVVYEPKK"
complement(13574. .13579)
complement(14226. .14231)
complement(14237. .14929)
/codon_start=1
/transl_table=11
/product="px01-12"
/protein_id="AAD32316.1"
/db_xref="GI:4894228"
/transl_table="MLKILICISFISFLIFFVIPTQDITMTKIVFIENGIGISI
PTIVAMQOLKOTIKLDFIDIKEMKETTKTEITNFVGLLKLOVYIGDILKRLSHYIN
RESMHGLYREFRDMKYNAYIIPINVAEKLIDVVASLLELEPHLOKIVLSY
VFTCAEYLNALPSIGSGSEKXKRVKLLSRNIQKNVILSAGFEKVFKEVEETYTE
FTFLVQKNRG"
complement(15040. .19002)
/note="erythrocyte invasion and possible binding protein;
Rhotry protein (2401 aa), Plasmodium yoelii (U36927);
520/1194 positive aa (43%)"
/codon_start=1
/transl_table=11
/product="px01-13"
/protein_id="AAD32317.1"
/db_xref="GI:4894229"
/transl_table="MAKKEIPDIKYEDIKAAKTGNLIIFFVAGISMLGLPSWKG
YAQELSDALIEKGVIDSLEYKLEKEDPKKILTKMLFEKELIPKSNELPFKPK
BEYKAIYKELYSNNAIYITWYDECLDVITTPQSDNEKVSNDUSENKLUTOSSET

```

```

Query Match 22.6%; Score 282.8; DB 1; Length 181654;
Best Local Similarity 57.8%; Pred. No. 5e-61;
Matches 503; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

QY 383 TCATTGTTTCACGTCAGAGAAAATCAAGAAAGCGTCATGCACAGCGCTGGCCCTTG 442
DB 111615 TAATGTCATCGTTTACCTAGAAATTCGGGAAAGACATGCTCAATTATGGGCTTTA 111674

QY 443 AAAGATCAGAGCGTGTCTTTTGTACCTGACTCAGATATATATCTACCTCGATG 502
DB 111675 AACGCAACACAGAGATGCTATTGTTACCATTCATTACAGATGCTGTTGCCAAATG 111734

QY 503 CTTTACAGAGTGTGTTAAACACCTTTAATGACCAACTGTTTTTGTGCGAGCGGTAC 562
DB 111735 CTGTTAGAGAGTATTGAAACCTTTAATGATGAAAAGTAATGCCACAACGTGTCAG 111794

QY 563 TTAATGTCAGAAATACAGAAACCAATCTCTTACACGGTTCACAGATATTCGCTATGATA 622
DB 111795 TGAACATTCGTAATAGAAATGATATTTAATAACAAAATAATTGATATGCGTTATGACA 111854

```

1. 181677

/organism="Bacillus anthracis str. A2012"

/mol\_type="genomic DNA"

/strain="A2012"

/db\_xref="taxon:191218"

/plasmid="pX01"

/note="Florida strain"

462..554

/gene="BXA0001"

CDS

462..554

/gene="BXA0001"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AA025958.1"

/db\_xref="GI:20520076"

/translation="MVARIHLIKFKVIVKSKRHSNECLFLMTK"

complement(996..1109)

/gene="BXA0002"

CDS

complement(996..1109)

/gene="BXA0002"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AA025959.1"

/db\_xref="GI:20520077"

/translation="MLVIGFLVISIVAIRLNMYYKLGKNKKDFVMKNG"

complement(1127..1765)

/gene="BXA0003"

CDS

complement(1127..1765)

/gene="BXA0003"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein, (pX01-01)"

/protein\_id="AA025960.1"

/db\_xref="GI:20520078"

/translation="MEVILFELILAVLIPLNSVKKVHPKWKAGKLVKRMLSKL  
DPKSYVLHNVVTVYEGDGTIDHVIAGTGVFVETKMYEGWYIGSERAAKRTGIGI  
FRKSSQPFQFNQKHKAEMLEIQQLPCISMAAFPKCSLKRVNVHSKEKHVLY  
NDLQKESITSDQVNDVEVOHVIYFTILKANIMDKIKERKKVYLHNKFAQ"

complement(1782..2165)

/gene="BXA0004"

CDS

complement(1782..2165)

/gene="BXA0004"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AA025961.1"

/db\_xref="GI:20520079"

/translation="MNRNRGVIALRNSISGLMTTIGVLSITSSVVYLLIGCFIL  
SLKVGMLGVKVKVEKRLPTWANGVLEKPVNRRVTFPQPRGLLVSLALGFVEMVVK  
GLKISYFIIGLTETHVFLMRNARM"

2291..2875

/gene="BXA0005"

CDS

2291..2875

/gene="BXA0005"

/note="identified by Glimmer2; putative"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AA025962.1"

/db\_xref="GI:20520080"

/translation="MLKLVYNTGPEGDSAFNSFTCKDCVYVLSKSDSCSIQLAA  
DSNNPLVPYECGNEFEYEIGTDDPLNSHIIHQQLREKAIQLESYPAQISDPIDPIKO  
AIVGWSVNGCWINDSKKFALAEQVFPFGDNNSKKYASFPFLDHDGAKEGIRIYH  
CWYVNEKGIGIEYHLLIGSVITMKIAEGVKKLGFNT"

2895..3473

/gene="BXA0006"

CDS

2895..3473

/gene="BXA0006"

```

CDS
    /gene="BXA0004"
    complement(1782..2165)
    /gene="BXA0004"
    /note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA025961.1"
/db_xref="GI:20520079"
/translation="MNRREYIALRNSISQWLLTCIVIGLSYFSSVYVLLIGCRIL
SLKYGVLGVKKVKRKPTWANFGVIETKPVRNRRTRPFRGGLLVSLALGFEMVIK
GLCISTFLIGLETHVFRLMRNRM"
2291..2875
/gene="BXA0005"
2291..2875
/gene="BXA0005"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA025962.1"
/db_xref="GI:20520080"
/translation="MLKIWNNTDPEGDSAFSNFTCKDCVYLKSQSDCSQLQAA
DSNNPLAVPGCCNFETYEITGDTPLNHSIHQOLRREKAQILAESYPQAQSIDPIHKO
AIHWGSVNYGCWLINDSKRFALAEGVPFGDNMSKKYASFPFLDHGAKEGIRIHY
CWYNKGIGETHLLDIGVTIMKTAEGVKKLGFTY"
2895..3473
/gene="BXA0006"
2895..3473

```

```

/ gene="BXA0006"
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein, (pX01-02)"
/ protein_id="AAM25963.1"
/ db_xref="GI:20520081"
/ translation="MLEHTILSLVOLPLIVSTGIYILLELTETWKWVYEGIF
AIIFLTRQVONKSIISLQHKWFSKINIFLVLFIASIPITLKLAFILOM
TVILFLEYCLYIDRLPIHNLRYKKEICDKQKLERIPNNEHGTHTIKTRDEQ
YFIYTAVLGSLTSPISEEVEEYNNLMGR"
complement(3524..4111)
/ gene="BXA0007"
complement(3524..4111)
/ gene="BXA0007"
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein, (pX01-03)"
/ protein_id="AAM25964.1"
/ db_xref="GI:20520082"
/ translation="MILLIPLQAAQASQVDFENKLGNDVSSFFKGLNNRTKITEK
DSFGSIQIEGLYAPAFKGVNIFISFLVGVIVVMVALITKNQWKMATGSKI
FSPISMSRLSVFLPSSDFSNFIPTDLDLFLKATLIPTPLIAGIRLRRIHE
STKQPEYRWANRVITGSCILCLMLAVLAPWLFAM"
complement(4159..4788)
/ gene="BXA0008"
complement(4159..4788)
/ gene="BXA0008"
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein, (pX01-04)"
/ protein_id="AAM25965.1"
/ db_xref="GI:20520083"
/ translation="MNINSISNNVKELOTVREKKQVQLRVSKSSNHYMIFP
ERSEKYNTELDEYINLSVSVYIDRLSQEIEKONHSEFMCTEVEDVIDDE
THENTTVKPTANKVFKLGGNLSDELHITVLNNINRIEIEFINENETY
VAVEKEELANAGTKANIYYAADLYAAEGNEIKLASINPKLAGDV"
complement(5124..5837)
/ gene="BXA0009"
complement(5124..5837)
/ gene="BXA0009"
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein, (pX01-05)"
/ protein_id="AAM25966.1"
/ db_xref="GI:20520084"
/ translation="MFTKIFIGFYVIVVCSKKFLREAKKKEMENKPLRPVY
WCPRENKPIGDEORTVPTKTEILKDVASIYDLVATSVHKNLGEVSDRLKHEDIE
EVEYAEDEPHPTPEAVYDIHPDSDEIDPVELLEYEDMPQOEIVSIPPETEPFVQ
QRVADYKMDREVRTKRVLEQEGGLRVINMSNKRYWIHTEGEILPENOLLYMKLEY
DSPSYTIVWSNRQIGNI"
complement(5854..6561)
/ gene="BXA0010"
complement(5854..6561)
/ gene="BXA0010"
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein, (pX01-06)"
/ protein_id="AAM25967.1"
/ db_xref="GI:20520085"
/ translation="MIFYQLPYSIFALNQARRHCRIGTKQCRFFYLVYRETFQOM
AQILAKKNAAEAMNGEATSGLNMLGDTGLQTLILIONIKSGNOLKSTEWKAA
SESRKLVANINVTIPVFESIEIOFTNWKEVLTQFMDVAFMKAOQVOKAEVCKKE
QVPGVYKQNVLEVDGEIEDIIRALTTLNTDKRMIVOKQDITFWDVIVVEVDSKKGS
KOKNKIIDQYFEDLFG"
complement(6678..8486)
/ gene="BXA0011"
complement(6678..8486)
/ gene="BXA0011"

```

```

/ note="identified by match to PFAM protein family HMM
PF00078"
/ codon_start=1
/ transl_table=11
/ product="reverse transcriptase/endonuclease protein,
(px01-07)"
/ protein_id="AAM25968.1"
/ db_xref="GI:20520086"
/ translation="MNTVMKSAVSQVOTWNISQIOWKEIENYVKLRQIRIARAFQFN
ARKKQRLMRURKANLLSIRKVTQIQNGKRTAGIDGTTTIPEDRVKFLHLLKY
SVRNKAFVKKRAYIPKNGKKRPLGIPVKDRIFQNMVKNALPEQWECRFESMSYGF
RPKRSADMANLFLKLSRGTNRNFIWFGDFQCFDNLNHEHILSCIEGFYSNAIQ
WLNAGCIDNKTKTFTGTGTPOGTTISPELLANIALHGMKEGLVYVHFPRGAGMLYPD
SIGIVRYADDVIVCNSEKAEASNYAKLQPYLDKRGKLAEEKRVVHITDGFDFGF
NFRQYPTKEGSQLFKPSNQSVYKAEKSEIFKSHRGRSISGLIRKLNPVITGIANY
WSPVAKYIIGDIDSTIVHKVYHLLKYKNSPDDANLKRYFAKDERIKFNRENTSKKLAK
PDNKNOLIRMTWTFIVRHVLIYKNSPDDANLKRYFAKDERIKFNRENTSKKLAK
KTKYKRCVNNLSVGEESNHIKVPVIGGKDEYDNLLELLHSCCHKQHALLWEYGN
GKQLPKVQAYLAKSQIPINSKKAIVRTMLNTEKKFY"
complement(8991..9083)
/ gene="BXA0012"
complement(8991..9083)
/ gene="BXA0012"
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AAM25969.1"

Query Match      22.6%; Score 282.8; DB 1; Length 181677;
Best Local Similarity 57.8%; Pred. No. 5e-61;
Matches 503; Conservative 0; Mismatches 367; Indels 0; Gaps 0;

QY 383 TCATGTTTCATCGTCAGAGAAAATCAAGAAAGCGTCATGCACAGCGCTGGGCTTTG 442
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111666 TAATGTCATCGTTTACCTAGAATTCGCGGAAAACAGATGCTCAATTATGGGTTTTA 111725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 443 AAGATCAGAGCGTGTGCTTTTGTGACCGTTCAGTACATCTTATCTACCCGTGATG 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111726 AACGGACAACAGCAGATGCTATTGTTACCATGATTGATGAGTGTGATTGTTCCCAATG 111795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 503 CTTTAGAGGAGTGTGTAACCACTTTAATGACCAACTGTTTTGCTGGAGCGGTCAAC 562
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111786 CTGTTAGAGAGTTTATTGAAACCTTTAATGATGAAAAAATATGGCCACAACGTGTCACG 111845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 563 TTAATGTCAGAAATAGACAAACCAATCTCTTAACAGCTTGACAGATATTCCTGATGA 622
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111846 TGAACATTCGTATAGAAATGATATTTATTACAAACTAATTGATATGCGTTATGACA 111905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 623 ATGCTTTTGGCGTTGAACGAGTGCCTGATCCCAATCCGTTACAGGTAAATATCTTGTCTCAG 682
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111906 ATGCGTTCCGTGGAGCGTGAGCAGACAGTCCGTAACAGGAAATGTTCTTGTGTTAGTG 111965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 683 GTCCCGCTTAGCGTTACACAGCCGAGTGGTGTTCCTAACATAGATAGATACATCAAC 742
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111966 GCGCTTAAAGTTGTGTTATCGTAGAAGTAATAACTGAAATTTAGAACATTTAGGAAGTC 112025
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 743 AGACCTTCCTGGGTATTCTCTGTAAGTATTGGTGATGACAGCTGCTTGACCAACTATGCA 802
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112026 AGATGTTCTTGGTGAGAGGTGTCAGTTTGGAGATGATAGATGCTAATAATATGCTA 112085
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 803 CTGATTTAGGAAAGACTGTTTATCAATCCACTGCTAAATGTTATTACAGATGTTCTGCTGACA 862
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112086 TTTTGAAGGGGAAACAGCTTTATCAATCCACTGCTCGATGTTATTACTGATGCTCCAACTA 112145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 863 AGATGCTCTACTTACTTGRAGCAGCAACACCGCTGGAACAAGTCTCTTCTTTAGAGAGTCCA 922
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112146 CATTAACAACATTTCTTAACAGCAACACAGTGTGGAACAAGTCAATTTTTCAGCAAGTT 112205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 923 TTATTTCTGTTAAGAAAATCATGAACAATCCTTTTGTAGCCCTATGACCATATCTTGGG 982
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 112206 TAATTCACITGGCATTGGTATGAAAAAACCAAAATGTTCTTGTGTTGGACAATTTTCGAAA 112265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

983	QY	TGCTATGTTTTATGAGCTGCTGTTTATTCGTGGTGGATTCCTTTGTFAGGCAATGTCAGAG	1042
112266	Db	TATCGTTATGAGATTTTATTTGGGTTTCCCTACTCTCAAGTATTTATCTCAAGCAAGTC	112325
1043	QY	AATTGTGATGGCTCAGGGTTTTAGCCTTTCTCGGTGATTATCTTCAATGTTGCCCTGTGTC	1102
112326	Db	ATGTAGGGTTAAATTTGGCTGTTTATTTATTTGGGTTATTTTCATTGCTGTATGCTTA	112385
1103	QY	GGAACTCATTTACATGCTTAAAGCACCGCTGCTCCTCTGTGTTATCTCCGTTTTATGGGG	1162
112386	Db	GAAATGTATTTTATCTATTATAAACATCCGCTACTTCTTACTGGCGCATTAATATGAA	112445
1163	QY	TGCTCATTTGTTTGCTACAGCGCTTGAAATATATATCTCTTTTACTATTAGAAATG	1222
112446	Db	TTCTCCATGTTATGACATATATACCTATACGGTTTTATGCTTTTACTAACTATTAAATCTTA	112505
1223	QY	CTGACTGTGGGACACGCTAAAAATATTATAT	1252
112506	Db	ATGGGTTGGGACACCGCTTAATTACAGTAAT	112535

Search completed: October 2, 2003, 10:30:04  
Job time : 4792 secs

PD 14-MAY-1999.

XX 30-OCT-1998; 98WO-US23153.  
 XX PF  
 XX 26-OCT-1998; 98US-0178951.  
 PR 31-OCT-1997; 97US-0064435.  
 XX  
 XX (OKLA ) UNIV OKLAHOMA STATE.  
 XX PA  
 XX PI DeAngelis P, Kumari K, Weigel PH;  
 XX WPI; 1999-337486/28.  
 DR P-PSDB; AAY06206.  
 XX  
 XX Nucleic acid encoding hyaluronate synthase for production of  
 PT hyaluronic acid with controlled molecular weight and targeting  
 PT specificity  
 XX  
 PS Claim 3; Page 118; 125pp; English.  
 XX  
 CC This DNA sequence codes for a hyaluronate synthase, termed seHAS  
 CC (see AAY06206), of group C Streptococcus equisimilis. To isolate  
 CC the DNA, genomic DNA was subjected to PCR amplification using  
 CC degenerate primers (see AAY58847-52) based on known related  
 CC sequences, and PCR products were used as probes to screen a DNA  
 CC library. Expression of the seHAS gene correlates with virulence  
 CC of streptococcal Group A and Group C strains, by providing a means  
 CC of escaping phagocytosis and immune surveillance. The invention  
 CC also provides recombinant vectors containing the isolated DNA, and  
 CC prokaryotic or eukaryotic host cells which produce seHAS and its  
 CC hyaluronic acid product, particularly a product with modified  
 CC structure or molecular size. The hyaluronic acid produced this way  
 CC is purer than that produced by conventional methods. The seHAS  
 CC gene also provides a new probe to assess the potential of bacterial  
 CC specimens to produce hyaluronic acid.  
 XX  
 SQ Sequence 1254 BP; 335 A; 240 C; 261 G; 418 T; 0 other;

Query Match 100.0%; Score 1254; DB 20; Length 1254;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGACATTAATAAAGCTCATACCTGTGTGGCTTTAGTATTTTGGTACTGTG 60  
 DB 1 ATGAGACATTAATAAAGCTCATACCTGTGTGGCTTTAGTATTTTGGTACTGTG 60  
 QY 61 ATTACGTCATATCTCTTTGGTCTAAAGAAAGCTGTCAATTTATGGCTTTG 120  
 DB 61 ATTACGTCATATCTCTTTGGTCTAAAGAAAGCTGTCAATTTATGGCTTTG 120  
 QY 121 CTGATAGCTTACCTATTAGTCAAAATGCTGTTATCCCTTTTACAGCCATTAAAGGA 180  
 DB 121 CTGATAGCTTACCTATTAGTCAAAATGCTGTTATCCCTTTTACAGCCATTAAAGGA 180  
 QY 181 AGGCTGGGCAATATAAGTTGACGACATTTCCCTTTATACCAAGATGCTGACTCA 240  
 DB 181 AGGCTGGGCAATATAAGTTGACGACATTTCCCTTTATACCAAGATGCTGACTCA 240  
 QY 241 TTGCTAGACCTTAAAAGTTTTCAGCAGCAAACTATCCCTAGCAGAAATTTATGTT 300  
 DB 241 TTGCTAGACCTTAAAAGTTTTCAGCAGCAAACTATCCCTAGCAGAAATTTATGTT 300  
 QY 301 GTTACCATGGAAGTCTGATGACAGCAGTATTAGCCATTAGCAGTATGTCGTTGAC 360  
 DB 301 GTTACCATGGAAGTCTGATGACAGCAGTATTAGCCATTAGCAGTATGTCGTTGAC 360  
 QY 361 ACTGGTACCTATCAGCAATGTCATTTGTCATCGGTGACAGAAAATCAAGGAAAGCGT 420  
 DB 361 ACTGGTACCTATCAGCAATGTCATTTGTCATCGGTGACAGAAAATCAAGGAAAGCGT 420  
 QY 421 CATGCAGACGCTGGGCTTTGAAAGATCAGACGCTGATGCTCTTTTGGCCCTTGACTCA 480  
 DB 421 CATGCAGACGCTGGGCTTTGAAAGATCAGACGCTGATGCTCTTTTGGCCCTTGACTCA 480

QY 481 GATACCTATTATCTACCTGATGCTTTAGAGAGTGTGTTAAAAACCTTTAATGACCAACT 540  
 DB 481 GATACCTATTATCTACCTGATGCTTTAGAGAGTGTGTTAAAAACCTTTAATGACCAACT 540  
 QY 541 GTTTTGTGCGAGCGGTCACTTAACTGTCAGAAATAGACAAACCAATCTCTTAACAGCG 600  
 DB 541 GTTTTGTGCGAGCGGTCACTTAACTGTCAGAAATAGACAAACCAATCTCTTAACAGCG 600  
 QY 601 TTGACAGATATTCGCTATGATAATGCTTTTGGCGTTGAACGAGCTGCCCAATCCGTTACA 660  
 DB 601 TTGACAGATATTCGCTATGATAATGCTTTTGGCGTTGAACGAGCTGCCCAATCCGTTACA 660  
 QY 661 GGTAAATATCCCTGTTTGTCTCAGTCCGCTTAGCGTTTACAGACGGGAGTGTGTTGCT 720  
 DB 661 GGTAAATATCCCTGTTTGTCTCAGTCCGCTTAGCGTTTACAGACGGGAGTGTGTTGCT 720  
 QY 721 AACATAGATAGATACATCAACACGACCTTCTCTGGTATTCTCTGAAGTATTGTTGATGAC 780  
 DB 721 AACATAGATAGATACATCAACACGACCTTCTCTGGTATTCTCTGAAGTATTGTTGATGAC 780  
 QY 781 AGTGCTTGCACCACTATGCAACTGATTTAGGAAAGACTGTTTATCAATCCACTGCTAAA 840  
 DB 781 AGTGCTTGCACCACTATGCAACTGATTTAGGAAAGACTGTTTATCAATCCACTGCTAAA 840  
 QY 841 TGTATTACAGATGTTCTCTGACAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAC 900  
 DB 841 TGTATTACAGATGTTCTCTGACAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAC 900  
 QY 901 AAGTCTCTTTTAGAGAGTCCATATTTCCTGTTAAGAAATCATGAACAATCCCTTTGTA 960  
 DB 901 AAGTCTCTTTTAGAGAGTCCATATTTCCTGTTAAGAAATCATGAACAATCCCTTTGTA 960  
 QY 961 GCCATATGACCATCTAGAGTGTCTATGTTTATGCTGTTTATCTGTTGGTGGAT 1020  
 DB 961 GCCATATGACCATCTAGAGTGTCTATGTTTATGCTGTTTATCTGTTGGTGGAT 1020  
 QY 1021 TCTTTTAGGCAATGTGACAGAAATTTGATTTGGCTCAGGGTTTATAGCTTTCTGTTGAT 1080  
 DB 1021 TCTTTTAGGCAATGTGACAGAAATTTGATTTGGCTCAGGGTTTATAGCTTTCTGTTGAT 1080  
 QY 1081 AUCTCATGTTGCCCTGTGCGAACATTCATTACATGCTTTAGCAGCCGCTGCTTC 1140  
 DB 1081 AUCTCATGTTGCCCTGTGCGAACATTCATTACATGCTTTAGCAGCCGCTGCTTC 1140  
 QY 1141 TCTTATCTCCGTTTATGCGGTGCTGATTTGTTCTTACAGCCCTTGAATATAT 1200  
 DB 1141 TCTTATCTCCGTTTATGCGGTGCTGATTTGTTCTTACAGCCCTTGAATATAT 1200  
 QY 1201 TCTCTTTTACTATTAGAAATGCTGACTGGGGAACAGCTAAATAATTTATATAA 1254  
 DB 1201 TCTCTTTTACTATTAGAAATGCTGACTGGGGAACAGCTAAATAATTTATATAA 1254

RESULT 2  
 ABS57951  
 ID ABS57951 standard; DNA; 1512 BP.  
 XX  
 AC ABS57951;  
 XX  
 DT 04-FEB-2003 (first entry)  
 XX  
 DE S. pyogenes hyaluronate (HA) synthase gene, hasA.  
 XX  
 KW Gene; ds; hyaluronate synthase; HasA; HA; infection; phagocytic cell;  
 KW polysaccharide capsule; hyaluronic acid; nonimmunogenic; macrophage;  
 KW pharyngitis; impetigo; deep tissue infection; rheumatic fever;  
 KW toxic shock syndrome; purity; cosmetic; eye surgery; viscosity;  
 KW biocompatibility; vitreous fluid.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 XX Key Location/Qualifiers  
 FT misc\_signal 193..264

```

FT      /*tag= a
FT      /note= "Possible start of the HasA open reading utilising
FT      a GIG start codon"
FT      238..264
FT      misc_signal
FT      /*tag= b
FT      /note= "Possible start of the HasA open reading utilising
FT      a TTG start codon"
FT      250..264
FT      misc_signal
FT      /*tag= c
FT      /note= "Possible start of the HasA open reading utilising
FT      a TTG start codon"
FT      265..1452
FT      CDS
FT      /*tag= d
FT      /product= "Mature HasA protein"
FT      1488..1508
FT      CDS
FT      /*tag= e
FT      /product= "HasB protein"
FT      /partial
FT      /note= "No stop codon shown"
XX
XX      US6455304-B1.
XX
XX      24-SEP-2002.
XX
XX      03-SEP-1998; 98US-0146893.
XX
XX      01-JUL-1994; 94US-0270581.
XX
XX      (UYOK-) UNIV OKLAHOMA.
XX
XX      Weigel PH, DeAngelis PL, Papaconstantinou J;
XX      WPI; 2003-066235/06.
XX      P-PSDB; ABG72066, ABG72067.
XX
XX      New purified nucleic acid segment encoding hyaluronate synthase, useful
XX      in recombinant DNA technology for the preparation of hyaluronic acid -
XX      Claim 2; Fig 7; 31pp; English.
XX
XX      The invention discloses a purified S. pyogenes nucleic acid encoding the
XX      hyaluronate (HA) synthase A (HasA). Streptococcal infections are a major
XX      health and economic problem and one reason for this is the ability of the
XX      bacteria to grow undetected by the body's phagocytic cells. They achieve
XX      this by coating themselves with polysaccharide capsules, such as
XX      hyaluronic acid capsules, which are generally nonimmunogenic and also
XX      prevent attachment of the bacteria to macrophages. Group A Streptococcus
XX      are responsible for numerous human diseases including pharyngitis,
XX      impetigo, deep tissue infections, rheumatic fever and toxic shock-like
XX      syndrome. The methods and compositions of the present invention are
XX      useful for recombinant DNA technology in the preparation of hyaluronic
XX      acid which has one or more improved properties, such as greater purity or
XX      ease of preparation, and produces larger amounts of relatively higher
XX      molecular weight and purity of hyaluronic acid than in prior art. High
XX      molecular weight HA has a wide variety of useful applications, ranging
XX      from cosmetics to eye surgery. In the case of the latter, its high
XX      viscosity and biocompatibility allows it to be a replacement for vitreous
XX      fluid. The sequence presented is the S. pyogenes HA synthase gene, hasA.
XX
XX      Sequence 1512 BP; 484 A; 227 C; 245 G; 556 T; 0 other;
XX
XX      Query Match      48.2%; Score 604.6; DB 25; Length 1512;
XX      Best Local Similarity 68.1%; Pred. No. 2.6e-156;
XX      Matches 848; Conservative 0; Mismatches 384; Indels 3; Gaps 1;
XX
XX      10 TTTAAACCCATCAACTGTGGCTTTAGTATTTTGGGTACTGTTGATTTACGTC 69
XX      || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      202 TTTAAACCAACCTTTAATGTTGTTTATCTCTTTATTTTATCTATCTCTGATTTATCTA 261
XX      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      70 AATGTTTATCTCTTTGCTGCTAAAGGAAGCTTGTCATTTATGCGCTTTTTCGCTGATAGCT 129
XX      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX      262 AATGATATCTATTGGAC---ATCAACTGTAGGATTTATGGAGTAATTAATTAACC 318

```

```

130 TACCTATTAGTCAAAATGCTCTTATCCTTTTTCACAGCCATTTTAAAGGAAGGGCTGG 189
131 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
319 TATCTAGTTATCAAACTGGATGATATCTTCTTTATGAGCCATTTAAGGAATCCATC 378
190 CAATATAAGGTGGAGCCATTTATCCCTCTTTATAAGCAAGATGCTGAGTCAATGCTAG 249
191 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
379 GACTATAAGTGTGCTGTGTAATCTCTTCTTATTAAGAGATGCGAGTCAATATTAGAA 438
250 ACCTTAAAGTGTTCAGACCAACCTTATCCCTAGCAGAAATTTATGTTGTTGACAT 309
251 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
439 ACCTTAAAGTGTGTTAGCAGACCTTATCCGTTATCAGAAATTTATATTGTTGATG 498
310 GGAAGTGTGATGAGACAGGATTAAGCGCATTTAGAGACTATGTCGTGACATGCTGAC 369
499 GGGAGTTCAACACAGATGCAATACAATTAATTCAGAGATGATAAATAGAGAAGTGG 558
370 CTATCAAGCAATGCTATGTTTCATCGGTACAGAAAATCAAGGAAGCGTCTACACAG 429
559 ATTTGCGAAAGCTTATCGTTCACCGTTCCTTCAATAAAGGAACGCCATGCTCAA 618
430 GCCTGGGCTTTGAAAGATCAGACGCTGATGCTTTTTCACCGTTTGACTCAGATCTAT 489
619 GCGTGGGCAITTTGAAGATCTGACGCTGACGTTTTTAAACGCTAGACTCAGATCTAT 678
490 ATCTACCTGATGCTTTTAGAGAGTGTGTTAAAAACCTTTAATGACCCAACTGTTTGT 549
679 ATCTATCCAAATGCTTTAGAAGAACTCTCTAAAAGCTTCAATGATGACAGATTTATG 738
550 GCGACGGTCACTTATGTCAGAAATAGACAAACCAATCTCTTAACACGCTTGACAGAT 609
739 GCAACAGGACATTTGAATGCTAGAAACAGACAAACTAATCTATTAAACGCACTTACAG 798
610 ATTGCGTATGATAATGCTTTTGGCGTTGAAGAGCTGCCCAATCCGTTTACAGGTAATATC 669
799 ATCGGTTACGATATGCTTTGGGGTGGAGCGTCTCTCAATCAATTAACAGGTAATAT 858
670 CTGTTGCTGAGTTCAGTTCAGGTTTACAGACGCGAGGTGGTTCCTTACATAGAT 729
859 TTAGTTTGTCTCAGGACCATTTAGTATTTATCGAGCTGAAGTATTTCTTAACTTAG 918
730 AGATATCAACACGACCTTCTCGGTGATTTCTCTAGTATTTGGTGATGACAGCTGCTG 789
919 CGCTATAAAATCAAACTTCTAGTGTTCCTGTTAGCAATGGGATGATGATGTTTGA 978
790 ACCAACTATGCACTGATTTAGGAAGAGCTGTTTATCAATCCACTGCTAAATGATTAACA 849
979 ACAATATGCTATTGATTTAGGACGCACTGCTCAACCAATCAACAGCTAGATGATACT 1038
850 GATGTTCTTGACAGATGCTCTTACTTGAAGCAGCAAAACCGCTGGAACAGTCCCTTC 909
1039 GATGATACCTTTCCAAATTTAAAGAGTTTATTTAAAGCAACAAATCGATGGAATAATCTTT 1098
910 TTTAGAGAGTCCATTTATTTCTGTTAAAGAAATCATGAACATCTCTTTGTAGCCCTATGG 969
1099 TTTAGAGATCTATTATTCTGTTTAAAGAAATCTTTCTATATCCCATCGTTGCTTATGG 1158
970 ACCATCTTGAAGGCTCTATGTTTATGATGCTTGTGTTTATCTGTTGGTGAATTTCTTGA 1029
1159 ACTATTTTGAAGTCTGTTATGTTTATGATGTTGATGTTGTCGCAATTTGGGAATCTTTGTTT 1218
1030 GGCAATGTCAGAGAATTTGATTTGCTCAGGTTTATAGCCCTTCTGGTATTTATCTCAT 1089
1219 AATCAAGCTATTCAATATAGACCTTATTAACCTTTTTCCTTTTATCCATCATCTTTATC 1278
1090 GTTCCTCTGTGCGAAGCAATTCATATGCTTAAGCAGCCCGCTGCTCTTCTTCTATCT 1149
1279 GTTCCTTATGCTGTAATGTTCAATATATGTTGTCAGAAATCTCTGCTAGTTTCTTATCT 1338
1150 CCGTTTATGGGCTGCTGCAATTTGTTTGTCTACAGCCCTTGAAATATATATCTCTTTTT 1209
1339 CCTCTGTGTAAGATATATACATTTGTTTGTCTTACAGCCCTTAAACTTATTTCTTTATGC 1398
1210 ACTATTAGAAATGCTGACTGGGACACGCTAAAAA 1244

```

Qy



```

Db 1225 CCTCTGATGGAATATACACTGTTGCTCTACACCCCTAAACCTTATCTTTATG 1284
QY 1210 ACTATTAGAAATGCTGACTGGGGAACAGTAAAAA 1244
Db 1285 ACCATTAAAAATACGGAATGGGGAACAGTAAAAA 1319

RESULT 4
ABL41013/c
ID ABL41013 standard; cDNA; 1653 BP.
XX AC
XX ABL41013;
XX 30-JUL-2002 (first entry)
XX Murine hyaluronan synthetase 2 (HAS2) cDNA fragment.
XX DE
XX Hyaluronan synthetase 2; HAS2; tumour; gene therapy; mouse; gene; ss.
XX KW
XX Mus sp.
XX OS
XX JP2002065280-A.
XX PD
XX 05-MAR-2002.
XX PF
XX 01-SEP-2000; 2000JP-0266260.
XX PR
XX 01-SEP-2000; 2000JP-0266260.
XX PA
XX (SEK ) SEIKAGAKU KOGYO CO LTD.
XX DR
XX WPI; 2002-397661/43.
XX PT
XX A DNA vector for inhibiting the formation of a malignant tumor
XX PS
XX Claim 1; Page 6; 9pp; Japanese.
XX CC
XX The invention relates to a DNA vector containing a 20 nucleotide sequence
XX of the 3'-end of a mouse hyaluronan synthetase 2 (HAS2) cDNA sequence,
XX or DNA hybridisable with a polynucleotide complementary to it under
XX stringent conditions, and which can inhibit the translation of a
XX hyaluronan synthetase 2 gene in which the transcription initiating point
XX is recombined 5' upstream of the DNA. The vector is used for inhibiting
XX the formation of a malignant tumour. The present sequence represents a
XX mouse HAS2 cDNA fragment.
XX SQ
XX Sequence 1653 BP; 471 A; 387 C; 353 G; 442 T; 0 other;

Query Match 4.9%; Score 61.6; DB 24; Length 1653;
Best Local Similarity 49.1%; Pred. No. 1.4e-06;
Matches 22; Conservative 0; Mismatches 224; Indels 6; Gaps 2;

QY 474 TGACACGATATATATACCTACCCGATGCTTTAGAGAGGTGTTAAACCTT---TAA 530
Db 1024 TGACACGATATATGCTGACCCCTCATCTGTGGAGATGTTGAGGCTTAGAGGA 965
QY 531 TGACCAACACTGTTTGTGCGAGCGGTGACCTTAATGTCAGAAATAGACAAACCAATCT 590
Db 964 AGACCTATGTTGGAGGAGATGTCAGATTTTAAACAGTATGATTCG 905
QY 591 CTTAACAGCTTGACAGATATGCGTATGCTATGCTTTTGGCGTTGACAGAGCTGCCA 650
Db 904 GATCTCTCTCCAGCAGCTGAGATGCTGATGCTTTTAAATAGAAAGGCGCTGCCA 845
QY 651 ATCCGTTACAGTAATTCCTTTGCTCAGGTCGCTACGTTTACAGAGCGAGGT 710
Db 844 GTCTTATTTGCGTGTCCAGTGCATAGCGGCTCTCTGGAATGTACAGAAACCTCTT 785
QY 711 GGTGTGTTCTACATAGATAGATACATCAACGACCTTCTCGGGTATTCCTGTAGTAT 770
Db 784 GCTGCATGAATTTGGAGACAGTGGTACATCAGGAATTCATGGGTAACCAACGAGTTT 725

```

```

QY 771 TGGTGTAGCAGGTGCTTGACCACTATGCAACTGATTTAGGAAAGACTGTTTATCAATC 830
Db 724 TGGTGTAGCAGGTGCTTGACCACTATGCAACTGATTTAGGAAAGACTGTTTATCAATC 865
QY 831 CACTG---CTAAATGATATTACAGATGTTCTGACAAGATGCTTACTTACTTGAAGCAGCA 887
Db 664 GGCTCGGTCCAAGTGCCTTACTGAACTCCCATAGATATCTGAGTGGCTGAACACGCA 605
QY 888 AAACCGCTGGAACAGTCTTCTTTTAGAGAGT 919
Db 604 GACCCGTTGGAGCAAGTCTTACTTCGAGAGT 573

RESULT 5
ABZ76735
ID ABZ76735 standard; cDNA; 1659 BP.
XX AC
XX ABZ76735;
XX 01-APR-2000 (first entry)
XX Mouse hyaluronan synthase 2 encoding cDNA SEQ ID NO:2.
XX DE
XX Mouse; hyaluronan synthase; HAS; HAS2; enzyme; ophthalmological;
XX KW osteopathic; antiarthritic; gene therapy; angiogenesis inhibitor;
XX KW hyaluronic acid; dihydrazide; bioconjugate; dry eye syndrome; eye;
XX KW osteoarthritis; gene; ss.
XX OS
XX Mus sp.
XX Key
XX Location/Qualifiers
XX CDS
XX 1..1659
XX /tag- a
XX /product= "hyaluronan synthase 2 (HAS2)"
XX WO2003006068-A1.
XX 23-JAN-2003.
XX 10-JUL-2001; 2001WO-US21785.
XX 10-JUL-2001; 2001WO-US21785.
XX (CLEA-) CLEAR SOLUTIONS BIOTECH INC.
XX Dehazya P, Chen W;
XX WPI; 2003-221664/21.
XX P-PSDB; ABP96029.
XX Novel dihydrazide derivatized hyaluronic acid/nucleic acid bioconjugate
XX for treating dry eye syndrome, has derivatized hyaluronic acid
XX crosslinked to nucleic acid encoding protein with hyaluronan synthase
XX activity
XX Claim 19; Page 56-57; 62pp; English.
XX The present invention describes a dihydrazide derivatised hyaluronic acid
XX (HA)/nucleic acid (NA) bioconjugate (I), with derivatised HA linked to NA
XX encoding protein with hyaluronan synthase (HAS) activity, where NA has
XX sequence with 79-85% sequence identity (SI) to a sequence of 1752, 1659
XX or 1665 nucleotides (see ABZ76734 to ABZ76736) or a sequence encoding
XX proteins with 66-90% SI to a sequence of 583, 552 or 554 amino acids (see
XX ABP96028 to ABP96030). (I) has ophthalmological, osteopathic and
XX antiarthritic activities, and can be used in gene therapy and as an
XX inhibitor of angiogenesis, and as an inducer of expression of (HA) in
XX human corneal epithelial cell. (I) is useful for transfecting a cell of
XX an eye with nucleic acid, where the nucleic acid comprises (ABZ76734 to
XX ABZ76736), by contacting the cell with (I) comprising the nucleic acid.
XX (I) is useful for treating dry eye syndrome in an individual. (I) is
XX useful in gene therapy applications for the treatment of a variety of
XX medical conditions including dry eye syndrome or other medical conditions
XX where an increase in the production of (HA) in the eye would be

```

CC therapeutic (e.g., osteoarthritis of the articular joints). (1) is also  
 CC useful for inhibiting angiogenesis for the treatment of macular  
 CC degeneration or genes related to lipid biosynthesis that helps to restore  
 CC the lipid component of the tear film, and as reagents for in vitro  
 CC transformation of any cell, preferably a eukaryotic cell, more preferably  
 CC a human eye cell. The present sequence encodes mouse HAS2 which is used  
 CC in the exemplification of the present invention.

SQ Sequence 1659 BP; 445 A; 353 C; 388 G; 473 T; 0 other;

Query Match 4.9%; Score 61.6; DB 25; Length 1659;  
 Best Local Similarity 49.1%; Pred. No. 1.4e-06;  
 Matches 222; Conservative 0; Mismatches 224; Indels 6; Gaps 2;

QY 474 TGACTCAGATCTTAATCTACCTGATGCTTTAGAGGAGTTGTTAAACCTT---TAA 530  
 DB 633 TGACTCAGATCTTAATCTACCTGATGCTTTAGAGGAGTTGTTAAACCTT---TAA 530  
 QY 531 TGACCAACTGTTTGTGCGAGCGGTCACTTAAATGTCAGAAATAGACAAACCAATCT 590  
 DB 693 AGACCCATGTTGGAGGTGTTGGAGGAGATGTCAGATTTTAAACAGTATGATTCCTG 752  
 QY 591 CTTAACAGCGTTGACAGATATTCGCTATGATTAATGCTTTGGCGTTGAACGAGCTGCCA 650  
 DB 753 GATCTCTCTCTCAGCAGCGTGAGATACTGGATGGCTTTTAAATATAGAAAGGCGCTGCCA 812  
 QY 651 ATCCGTTACAGGTAATATCTTGTCTCAGGTCCGCTTAGCGTTTACAGACGCGAGGT 710  
 DB 813 GTCCTATTGTTGGCTGTCAGAGTCATGACCGGTCTCTGGAAATGTCAGAAATCTCTT 872  
 QY 711 GGTGTTCTTAAATAGATAGATCATCAACAGACCTTCTGGGTATTCCTGTAAGTAT 770  
 DB 873 GCTGATGAATTTTGGAGAGCTGTCAATCAGGAATTCATGGTTAAACCAATGCAAGTT 932  
 QY 771 TGGTGATCAGAGTGTTCACCACTATGCAACTGATTTAGGAAGACTGTTTATCAATC 830  
 DB 933 TGGTGACAGCAGCAGCTTACCAAGGCTGTTGAGTCTGGGCTATGCAACTAAATACAC 992  
 QY 831 CACTG---CTAAATGTTATACAGATGTTCTGACAGATGCTACTTACTTGTGAACGACGA 887  
 DB 993 GCGTCGTCCTCAAGTGCCTTACTGAACTCCCATAGATATCTGAGATGGCTGAACGACGA 1052  
 QY 888 AAACCGCTGGAACAAGTCTCTTTTAGAGAGT 919  
 DB 1053 GACCCGATGGAGCAAGTCTCTACTTCCGAGAGT 1084

## RESULT 6

AAZ88200

ID AAZ88200 standard; cDNA to mRNA; 2890 BP.

XX AC AAZ88200;

XX DT 27-APR-2000 (first entry)

XX DE Mouse hyaluronate synthase modified protein HAS2 encoding cDNA.

XX KW Mouse; murine; hyaluronate synthase; modification; HAS1; HAS2; HAS3;

XX KW research reagent; biochemical research; medical development; ds.

XX OS Mus sp.

XX FH Key Location/Qualifiers

XX FT CDS 508..2166

XX FT /\*tag= a

XX FT /product= "HAS2"

XX FT /note= "modified hyaluronate synthase protein"

XX PN JP2000004886-A.

XX PD 11-JAN-2000.

XX XX 24-JUN-1998; 98JP-0193788.

XX PF

XX PR 24-JUN-1998; 98JP-0193788.  
 XX PA (SEKG ) SEIKAGAKU KOGYO CO LTD.  
 XX DR WPI: 2000-140125/13.  
 XX DR P-PSDB; AAY68492.

XX A hyaluronate synthase modified protein - useful as a research reagent  
 PT for biochemical research and medical development  
 XX Claim 18; Page 16-19; 30pp; Japanese.

XX The present invention describes a recombinant protein which consists of  
 CC three continuous regions (N-terminal region, internal region and  
 CC C-terminal region) where one or two regions among the above three  
 CC regions is selected from the three hyaluronate synthase (HAS) modified  
 CC proteins HAS1, HAS2 and HAS3 (e.g. a protein made from HAS1 amino acids  
 CC 1 to 71, HAS2 amino acids 52 to 379 and HAS1 amino acids 410 to 583).  
 CC The protein is useful as a research reagent for biochemical research  
 CC and medical development. The invention provides a HAS modified protein  
 CC of increased or lowered activity. The present sequence encodes mouse  
 CC HAS2.

SQ Sequence 2890 BP; 826 A; 580 C; 631 G; 853 T; 0 other;

Query Match 4.9%; Score 61.6; DB 21; Length 2890;  
 Best Local Similarity 49.1%; Pred. No. 1.8e-06;  
 Matches 222; Conservative 0; Mismatches 224; Indels 6; Gaps 2;

QY 474 TGACTCAGATCTTAATCTACCTGATGCTTTAGAGGAGTTGTTAAACCTT---TAA 530  
 DB 1092 TGACTCAGATCTTAATCTACCTGATGCTTTAGAGGAGTTGTTAAACCTT---TAA 530  
 QY 531 TGACCAACTGTTTGTGCGAGCGGTCACTTAAATGTCAGAAATAGACAAACCAATCT 590  
 DB 1152 AGACCCATGTTGGAGGTGTTGGAGGAGATGTCAGATTTTAAACAGTATGATTCCTG 1211  
 QY 591 CTTAACAGCGTTGACAGATATTCGCTATGATTAATGCTTTGGCGTTGAACGAGCTGCCA 650  
 DB 1212 GATCTCTCTCTCAGCAGCGTGAGATACTGGATGGCTTTTAAATATAGAAAGGCGCTGCCA 1271  
 QY 651 ATCCGTTACAGGTAATATCTTGTGCTCAGGTCCGCTTAGCGTTTACAGACGCGAGGT 710  
 DB 1272 GTCCTATTGTTGGCTGTCAGAGTCATGACCGGTCTCTGGAAATGTCAGAACTCTT 1331  
 QY 711 GGTGTTCTTAAATAGATAGATCATCAACGACCTTCTGGGTATTCCTGTAAGTAT 770  
 DB 1332 GCTGCAATGAATTTGGAAGACTGGTACATCAATCAGGAATTCATGGTAAACCAATGCAAGTT 1391  
 QY 771 TGGTGATCAGAGTGTCTTACCAACTATGCAACTGATTTAGGAAGACTGTTTATCAATC 830  
 DB 1392 TGGTGACGACAGCAGCCTTACCAACAGCGGTGTTGAGTCTGGGCTATGCACTAAATACAC 1451  
 QY 831 CACTG---CTAAATGTTATACAGATGTTCTGACAGATGCTCTACTTACTTGAACGACGA 887  
 DB 1452 GGCTCGGTCCAAGTGCCTTACTGAACTCCCATAGATATCTGAGATGGCTGAACGACGA 1511  
 QY 888 AAACCGCTGGAACAAGTCTCTTTTAGAGAGT 919  
 DB 1512 GACCCGTTGGAGCAAGTCTCTACTTCCGAGAGT 1543

## RESULT 7

AAV18821

ID AAV18821 standard; cDNA; 2947 BP.

XX AC AAV18821;

XX DT 11-JUN-1998 (first entry)

XX DE Murine hyaluronan synthase-2 cDNA.

XX XX

50 1440 10010ACGACAGGCACCTTACCCATACAGGGTGTGTGAGTCTGGGCTATGCAACATAATATACAC 1493

```

Db      1220 AGACCCATATGTTGGAGGTGTTGGAGGAGATGTCAGATTTTAAACAAGTATGATTCCTG 1279
QY      591 CTTACACGGCTGACAGATATTCCTATGATATGCTTTTGGCGTTGAACGAGCTGCCCA 650
Db      1280 GATCTCCTCTCAGACGGTGAGATCTGGATGGCTTTTAAATATAGAAAGGCGCTGCCA 1339
QY      651 ATCCGTTACAGGTAATATCCTGTTTCTCAGTCGCGTATAGCGTTTACAGACGGAGGT 710
Db      1340 GTCATATTTGGCTGTGTCAGTGCATATAGCGGTCTCTGGAAATGTACAGAACTCGCTT 1399
QY      711 GGTGTTCTTACATAGATAGATATCAACACAGACCTTCTCGGTATTCCTGTAACTAT 770
Db      1400 GTCATCAATTTTGTGAAGACTGGTACATCAGGAATTCATGGTAACCAATGCACTT 1459
QY      771 TGGTGATGACAGCTGTTGACCAACTATGCACTGATTTAGAAAGACGTGTTATCAATC 830
Db      1460 TGGTGACAGACGCCACTTACCAACAGGGTGTGAGTCTGGGCTATGCAACTAAATACAC 1519
QY      831 CACTG---CTAATGTATTACAGATGTTCTCTGACAAGATGTCATCTACTTCTGAAGCAGCA 887
Db      1520 GGCTGGTCCAAGTGCCTTACTGAACTCCCATAGATATCTGAGATGCTGAGTGCCTGACCCAGCA 1579
QY      888 AAACCGTGAACAGTCTTCCTTTAGAGAGT 919
Db      1580 GACCGTGGAGCAAGTCTCTACTTCCGAGAGT 1611

```

## RESULT 9

```

AAAA14914
ID  AAA14914 standard; DNA; 7193 bp.
XX
AC  AAA14914;
XX
DT  08-AUG-2000 (first entry)
XX
DE  DNA encoding nodulation efficiency factors.
XX
KW  Nodulation efficiency factor; Sinorhizobium meliloti USDA 1170; nodJ;
KW  strain NRG 185; cut-leaf medic; nod gene; nodA; nodB; nodC; nodD; nodI;
KW  nodulation factor; nodulation efficiency; Medicago lacinata;
KW  nitrogen fixation; legume; ss.
XX
OS  Sinorhizobium meliloti.
XX

```

```

FH  Key Location/Qualifiers
FT  CDS complement (3..275)
FT  /tag= a
FT  /product= "nodB"
FT  /note= "encodes AAY84842"
FT  508..511
FT  /tag= b
FT  526..1116
FT  /tag= c
FT  /product= "nodA"
FT  /note= "encodes AAY84843"
FT  1100..1103
FT  /tag= d
FT  1113..1766
FT  /tag= e
FT  /product= "nodB"
FT  /note= "encodes AAY84844"
FT  1766..1769
FT  /tag= f
FT  1781..3046
FT  /tag= g
FT  /product= "nodC"
FT  /trans_except= (pos: 1961..1963, aa: Glu)
FT  /note= "encodes AAY84845"
FT  3070..3074
FT  /tag= h
FT  3092..4138
FT  /tag= i

```

```

FT  /product= "nodI"
FT  /note= "encodes AAY84846"
FT  4125..4127
FT  /tag= j
FT  4142..4930
FT  /tag= k
FT  /product= "nodJ"
FT  /note= "encodes AAY84847"
FT  5075..5079
FT  /tag= l
FT  5103..5321
FT  /tag= m
FT  /product= "ORF1 protein"
FT  /note= "encodes AAY84848"
FT  5353..5358
FT  /tag= n
FT  5368..5706
FT  /tag= o
FT  /product= "ORF2 protein"
FT  /trans_except= (pos: 5533..5535, aa: Gln)
FT  /note= "encodes AAY84849"
FT  complement (5866..6297)
FT  /tag= p
FT  /product= "ORF3 protein"
FT  /note= "encodes AAY84850"
FT  6294..6297
FT  /tag= q
FT  complement (6306..6563)
FT  /tag= r
FT  /product= "ORF4 protein"
FT  /note= "encodes AAY84851"
FT  6559..6564
FT  /tag= s
FT  6567..6572
FT  /tag= t
FT  complement (6613..7118)
FT  /tag= u
FT  /product= "ORF5 protein"
FT  /note= "encodes AAY84852"
FT  7113..7117
FT  /tag= v
XX
PN  WO200022138-A1.
XX
PD  20-APR-2000.
XX
PF  13-OCT-1999; 99WO-CA00955.
XX
PR  14-OCT-1998; 98US-0104162.
XX
PA  (AGRI-) AGRIC & AGRI-FOOD CANADA.
XX
PI  Barran LR, Bromfield ESP, Brown DCW;
XX
PI  WPI; 2000-317991/27.
DR  P-PSDB; AAY84842, AAY84843, AAY84844, AAY84845, AAY84846, AAY84847,
DR  AAY84848, AAY84849, AAY84850, AAY84851, AAY84852.
XX
PT  Isolated nodulation efficiency factor for improving nitrogen fixation
PT  in legumes comprising a portion of EcoRI/BamHI fragment of
PT  Sinorhizobium meliloti that confers the nodulating ability -
XX
PS  Claim 1; Fig 1; 47pp; English.
XX
CC  The present sequence encodes nodulation efficiency factors. The present
CC  sequence is a 7.2 kb EcoRI/BamHI fragment from Sinorhizobium meliloti
CC  USDA 1170, that gives S. meliloti strain NRG 185 the ability to nodulate
CC  at least 50% of inoculated Medicago lacinata (cut-leaf medic) plants
CC  within 10 days of inoculation. The fragment contains nod genes nodA,
CC  nodB, nodC, nodD, nodI, nodJ. The nodulation factors are used to
CC  increase the nodulation efficiency of Sinorhizobium for Medicago
CC  lacinata. The nodulation factors are also used to improve nitrogen
CC  fixation in legumes.

```

```

XX SQ Sequence 7193 BP; 1434 A; 2094 C; 2132 G; 1532 T; 1 other;
Query Match 4.5%; Score 56; DB 21; Length 7193;
Best local similarity 46.6%; Pred. No. 9.1e-05;
Matches 179; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 412 GGAAGGCTCATGACAGCGCTGGGCTTTGAAAGATCAGACGCTGATGCTCTTTTGACC 471
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 472 GTTGACTCAGATCTTATATCTACCCGATGCTTTAGAGAGTGTGTTAAACACCTTTAAT 531
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 532 GACCCAACTGTTTTCTCGCGAGCGGTCACCTTAATGTCAGAAATACAAACCAATCTC 591
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 592 TTAACAGCGTTCAGACATATTCGCTATGATAATGCTTTTGGCGTTGACGAGTGCACAA 651
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 652 TCCGTTACAGGTAATATCTTTGCTCAGGTCCGCTTAGCGTTTACAGCGGAGGTG 711
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 712 GTTGTCTTAACATAGATAGATACATCAACAGACCTTCCTGGGATTCCTGTAGTATT 771
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 772 GGTATGACAGGTGCTTGACCAAC 795
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2498 GGTGAGGATCGTCACTGACGATC 2521

RESULT 10
AAV30458/c
ID AAV30458 standard; DNA; 534720 BP.
XX
AC AAV30458;
XX
XX 14-OCT-1998 (first entry)
XX
XX Rhizobium species plasmid pNGR234a.
XX
XX Symbiosis; open reading frame; ORF; plasmid; vector; transportation;
XX degradation; metabolism; host range; nitrogen fixation; nodulation;
XX legume; plant; ds.
XX
XX Rhizobium sp.
XX
XX Location/Qualifiers
FH 417796..418671
FT CDS
FT
FT /*tag= a
FT /standard_name= "ORF K1"
FT /product= "oligopeptide permease"
FT /note= "homologous to the OppC gene"
FT 418673..419680
FT CDS
FT
FT /*tag= b
FT /standard_name= "ORF K2"
FT /product= "oligopeptide permease"
FT /note= "homologous to the OppD gene"
FT 419677..420738
FT CDS
FT
FT /*tag= c
FT /standard_name= "ORF K3"
FT /product= "oligopeptide permease"
FT /note= "homologous to the OppF gene"
FT 420774..422159
FT CDS
FT
FT /*tag= d
FT /standard_name= "ORF K4"
FT /product= "encapsulation-like protein"
FT /note= "homologous to the CapA gene"
FT

422628..424031
/*tag= e
/standard_name= "ORF K5"
/product= "aminotransferase-like protein"
/note= "homologous to the BioA gene"
424056..425594
/*tag= f
/standard_name= "ORF K6"
/product= "(semi)aldehyde dehydrogenase-like protein"
complement (426949..428028)
/*tag= g
/standard_name= "ORF K7"
/product= "transposase homologue"
/note= "homologous to the tnp gene"
428292..429623
/*tag= h
/standard_name= "ORF K8"
/product= "glutamate dehydrogenase-like protein"
/note= "homologous to the GUD1 gene"
complement (430538..431284)
/*tag= i
/standard_name= "ORF K9"
/product= "transposase homologue"
complement (431296..432840)
/*tag= j
/standard_name= "ORF K10"
/product= "transposase homologue"
/note= "homologous to the tnp gene"
complement (433880..434110)
/*tag= k
/standard_name= "ORF K11"
/product= "protein of unknown function"
/note= "homologous to the FixJ gene"
complement (434107..434433)
/*tag= l
/standard_name= "ORF K12"
/product= "protein of unknown function"
complement (434517..434711)
/*tag= m
/standard_name= "ORF K13"
/product= "ferroxidase/ferroxidase-like protein"
/note= "homologous to the FdxN gene"
complement (434753..436234)
/*tag= n
/standard_name= "ORF K14"
/gene= "nifB"
/product= "protein involved in FeMo co-factor biosynthesis"
complement (436460..438130)
/*tag= o
/standard_name= "ORF K15"
/gene= "nifA"
/product= "positive regulator of nif, fix and other genes"
complement (438297..438590)
/*tag= p
/standard_name= "ORF K16"
/gene= "fixX"
/product= "protein required for nitrogenase activity"
complement (438605..439912)
/*tag= q
/standard_name= "ORF K17"
/gene= "fixC"
/product= "protein required for nitrogenase activity"
complement (439923..441032)
/*tag= r
/standard_name= "ORF K18"
/gene= "fixB"
/product= "protein required for nitrogenase activity"
complement (441042..441899)
/*tag= s
/standard_name= "ORF K19"
/gene= "fixA"

```



PF 10-JUL-1997; 97WO-IB00950.  
 XX  
 PR 20-MAY-1997; 97GB-0010395.  
 PR 12-JUL-1996; 96EP-0730001.  
 XX  
 PA (MOLE-) INST MOLECULAR BIOTECHNOLOGY.  
 PA (BIOL-) LAB BIOLOGIE MOLECULAIRE PLANTES SUPERIE.  
 XX  
 PI Broughton WJ, Freiberg CB, Perret XP, Rosenthal A;  
 XX  
 DR WPI; 1998-110606/10.  
 XX  
 PT New isolated symbiotic plasmid from Rhizobium sp. NGR234 - used to  
 PT develop products for modifying plant characteristics, e.g. nitrogen  
 PT fixation, synthesis of compounds and stress response  
 XX  
 PS Claim 1; Fig 3; 220pp; English.  
 XX  
 CC This is the nucleotide sequence of the plasmid pNGR234a isolated from  
 CC Rhizobium sp. NGR234. Open reading frames (ORF) derivable from the  
 CC nucleotide sequence are claimed. The nucleotide sequences or ORFs can  
 CC be used e.g. in the transportation of compounds to and from an organism  
 CC which is a host to at least one of the nucleotide sequences, ORFs or  
 CC proteins, the degradation and/or metabolism of organic, inorganic,  
 CC natural or xenobiotic substances in a host organism or the modification  
 CC of the host range, nitrogen fixation abilities; for obtaining a synthetic  
 CC minimal set of ORFs required for functional Rhizobium-legume symbiosis,  
 CC especially for nodulation efficiency on host plants.  
 XX  
 SQ Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;  
 Query Match 4.3%; Score 53.8; DB 19; Length 536165;  
 Best Local Similarity 46.1%; Pred. No. 0.0022;  
 Matches 181; Conservative 0; Mismatches 212; Indels 0; Gaps 0;  
 QY 403 AAAATPCAAAGAGCGTCATGCACAGCGCTGGGCGCTTTGAAAGATCAGACGCTGATGTC 462  
 DB 156991 AAGAAGCGTCGGAAGCGCAAGCGCAGATTGCGGATACGGGAATCATCGGAGATTG 156932  
 QY 463 TTTTGACCGTGTACTACATCTATATCTACCCGTGCTTAGAGGAGTGTAAAA 522  
 DB 156931 GTGCTCAACGTGTACTCGGACGACGACCATGCGCGGAGCTAGTCACGAACTGCCCTG 156872  
 QY 523 ACCTTAAATGACCAACTGTTTGTGTCGACGCGGTGTCACCTTAATGTCAAGAAATAGACAA 582  
 DB 156871 AAGATGTACAGTCCGCGGTGCGCGCGATGGGTGCTGACGCGCCACACCGCAGC 156812  
 QY 583 ACCAATCTTTAACAGCGCTGACAGATATTCGCTATGATATGCTTTGGCGTGAACGA 642  
 DB 156811 GACATCGCTGACGCGGTGATGACATGAGTACTGGCTGCGCTGCAACGAGGAACGC 156752  
 QY 643 GCTGCCCAATCCGTTACAGGTAAATATCCTTGTGTCAGGTCCGCTTAGCGTTTACAGA 702  
 DB 156751 GCAGCACAGGTCGCTTTGAGCCGCTATGCTGTGTCGCGCGCTGTCGCTATGACGG 156692  
 QY 703 CGGAGGTTGGTGTCTTAAATAGATAGATACATCAACAGACCTTCCTGGGTATTCCT 762  
 DB 156691 CGGTCGCACTTCCTATTGCTGCTGATAAATACGAGACCACTGTTTCGAGGAGGCCA 156632  
 QY 763 GTAAGTATTGGTGATGACAGGTGCTTGACCAAC 795  
 DB 156631 ACAGACTTCGGGGAGACCGCCACCTCACAATC 156599  
 RESULT 12  
 AAV18822  
 ID AAV18822 standard; cdNA; 1665 BP.  
 XX  
 AC AAV18822;  
 XX  
 DT 11-JUN-1998 (first entry)  
 XX  
 DE Murine hyaluronan synthase-3 cdNA.

XX Murine; hyaluronan synthase-3; HS-3; hyaluronan;  
 KW hyaluronic acid; ss.  
 XX  
 OS Mus sp.  
 FH Key Location/Qualifiers  
 FT CDS 1..1665  
 FT /\*tag= a  
 FT /product= hyaluronan\_synthase-3  
 XX  
 PN W09800551-A2.  
 XX  
 PD 08-JAN-1998.  
 XX  
 PF 03-JUL-1997; 97WO-US11761.  
 XX  
 PR 05-MAR-1997; 97US-0812008.  
 PR 03-JUL-1996; 96US-0675499.  
 XX  
 PA (MAYO-) MAYO FOUNDATION.  
 XX  
 PI Augustine ML, McDonald JA, Spicer AP;  
 XX  
 DR WPI; 1998-086976/08.  
 DR P-PSDB; AAW50010.  
 XX  
 CC Nucleic acid encoding hyaluronan synthase-2 or -3 or their active  
 CC fragments - used for treating arthritis, in wound repair, for  
 CC delivering drugs etc.  
 XX  
 PS Claim 11; Pages 75-78; 115pp; English.  
 XX  
 CC The present sequence encodes murine hyaluronan synthase-3 (HS-3),  
 CC which can be used to alter hyaluronan, i.e. hyaluronic acid (HA),  
 CC production, or for in vitro synthesis of HA.  
 CC HA can be used to treat arthritis or perforated ear drums, protect  
 CC eyes during surgery, deliver drugs, prevent post-operative scarring  
 CC or adhesion formation, heal wounds and prevent restenosis following  
 CC balloon angioplasty. Modulation of HA production in vivo may be  
 CC useful in, e.g. Graves disease, mesothelioma, Wilm's tumour and  
 CC oedema associated with inflammation of lung and kidney, all of  
 CC which are associated with elevated levels of HA in tissues or  
 CC serum.  
 XX  
 SQ Sequence 1665 BP; 340 A; 453 C; 451 G; 421 T; 0 other;  
 Query Match 3.9%; Score 49.4; DB 19; Length 1665;  
 Best Local Similarity 49.2%; Pred. No. 0.0033;  
 Matches 159; Conservative 0; Mismatches 161; Indels 3; Gaps 1;  
 QY 600 CTTGACAGATATTCGCTATGATATGCTTTGGGCTTGAACGAGTGCCTACCGTTAC 659  
 DB 774 CTTGACAGTGTGAGGTACTGGATGGCTTCAAGCTGGAGCGGCTGCCAGTCTACTT 833  
 QY 660 AGGTAAATATCTTGTGTTGCTCAGGTCCGCTTACAGCTGTTACAGCGAGGTGTTGTC 719  
 DB 834 TGGCTGTGCAATCATATTAGTGGGCTTTGGCATGTACCGCACAGCTCTTCAGCA 893  
 QY 720 TAACATAGATAGATACATCAACAGACCTTCCTGGGTATTTCTGTGAAGTATGGTGATGA 779  
 DB 894 GTTCTGGAGGATTTGGTACCATCAGAAGTCTTAGGCGAGCAAGTGCAGCTTTGGGATCA 953  
 QY 780 CAGTGCTTGCACCACTATGCACTGATTTAGGAAA---GACTGTTTATCAATCCACTGC 836  
 DB 954 TCGGACCTTACCAACCGAGTCTCTGAGTCTTGGCTACCGGACTAAGTATACAGACGCTC 1013  
 QY 837 TAAATGTATTACAGATGTTCTGACAAGATGTCCTACTTCTTGAAGCAGCAAAACGCTG 896  
 DB 1014 TAGTGCCTCAGAGAGACCCCTAGTACTTCCGATGGCTCAATCAGCAAAACCGCTG 1073  
 QY 897 GAACAAGTCTCTTTTAGAGAGT 919  
 || ||||| ||||| |||||

500	CTTCACAGATATTCCTATGATATCTCTTTTCGGGTGGACGAGTGCCTCAATCCGGTTAC	659
QY		
774	CTCTGACAGTGTGAGGACTGGATGGCTTTCAACGTGGAGCGGCGCTGCCAGTCCCTACTT	833
Db		
660	AGGTAATATCTCTTGTTCCTCAGGTCGCGCTTAGCGTTTACAGACGCGAGTGTGGTTGCC	719
QY		
834	TGGCTGTGTCAATGATATAGTGGGCGCTTGGCGATGTACCCACAGACGCTCCTTCACGA	893
Db		
720	TAACTAGATAGATACATCAACAGACCTTCCTGGGTATTCCTGTAAAGTATGGTGATGA	779
QY		
894	GTTCTCGGAGGATGGTACCATCAGAAGTTCCTAGGACCAAGTGCAGCTTTGGGGATGA	953
Db		
780	CAGGTGCTTGACCAACTATCAACTGATTTAGAA--GACTGTTTATCAATCCACTGC	836
QY		

CC degeneration or genes related to lipid biosynthesis that helps to restore  
 CC the lipid component of the tear film, and as reagents for in vitro  
 CC transformation of any cell, preferably a eukaryotic cell, more preferably  
 CC a human eye cell. The present sequence encodes mouse HAS3 which is used  
 CC in the exemplification of the present invention.  
 XX  
 SQ Sequence 1665 BP; 340 A; 453 C; 451 G; 421 T; 0 other;

Query Match 3.9%; Score 49.4; DB 25; Length 1665;  
 Best Local Similarity 49.2%; Pred. No. 0.0033;  
 Matches 159; Conservative 0; Mismatches 161; Indels 3; Gaps 1;  
 QY 600 CTGACAGATATCGGTATGATGCTTTTGGCGTTGACGAGCTGCCCAATCCGTTAC 659  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 774 CTGAGCAGCTGAGGTACTGATGGCTTTCACGTTGGAGCGGCGCTGCCAGTCCCTACTT 833  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 660 AGGTAAATATCCTTTTGGCTCAGGTCGCTTAGCTTTACAGACGCGAGGTGTTGTTCC 719  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 834 TGGCTGTGCAATGATTAGTGGGCTTTTGGGCAATGACGCAACAGCCTCCTTCAGCA 893  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 720 TAACATAGATAGATACATCAACGAGCCTTCCCTGGGTATTCCTGTAGTATTGGTGATGA 779  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 894 GTTCTGGAGGATTGGTACCATCAGAAAGTTCTTAGCAGCAAGTGCAGCTTTGGGGATGA 953  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 780 CAGGTGCTTGACCAACTATGCAACTGATTAGGAAA--GACTGTTTATCAATCCACTGC 836  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 954 TCGGACCTTACCAACGAGTCTCTGAGTCTTGGCTACCGGACTAAGTATACAGCAGGCTC 1013  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 837 TAAATGATATACAGATGTTCCGACAGATGCTTACTTACTTTGAAGCAGCAAAACCGCTG 896  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1014 TAAGTGCTTCACAGAGACCCCACTAGGTACCTTCGATGGCTCAATCAGCAAAACCGCTG 1073  
 QY 897 GACAACTCCTTCTTAGAGAGT 919  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1074 GACAACTTACTTTCGGGAAT 1096  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15  
 AAA10595/C  
 ID AAA10595 standard; DNA; 6741 BP.  
 XX  
 AC AAA10595;  
 XX  
 DT 29-JUN-2000 (first entry)  
 XX  
 DE Gene encoding a subunit of cellulose synthase.  
 XX  
 KW Cellulose synthase; cellulose production; increase yield; ds.  
 XX  
 OS Vigna angularis.  
 XX  
 PN JP2000060568-A.  
 XX  
 PD 29-FEB-2000.  
 XX  
 PF 26-AUG-1998; 98JP-0239998.  
 XX  
 PR 26-AUG-1998; 98JP-0239998.  
 XX  
 PA (MIZUO) MIZUNO K.  
 PA (OJIP) OJI PAPER CO.  
 XX  
 DR WPI; 2000-342371/30.  
 DR P-PSDB; AAY85180.  
 XX  
 PT A gene encoding a cellulose synthetic equipment - for the improvement  
 PT in the amount of cellulose synthesised in a plant body  
 XX  
 PS Claim 2; Page 22-31; 32pp; Japanese.  
 XX  
 CC This sequence represents a gene encoding a subunit of the cellulose  
 CC synthase complex of Vigna angularis. The invention relates to subunits of  
 CC cellulose synthetic equipment, that can be used to increase the amount of

CC cellulose synthesised by a plant. The proteins and genes encoding them  
 CC can also be used to improve the properties of the cellulose being  
 CC produced by a plant.  
 XX  
 SQ Sequence 6741 BP; 1712 A; 870 C; 1468 G; 1374 T; 1317 other;

Query Match 3.8%; Score 47.8; DB 21; Length 6741;  
 Best Local Similarity 18.4%; Pred. No. 0.016;  
 Matches 114; Conservative 201; Mismatches 301; Indels 2; Gaps 1;  
 QY 600 CTTGACAGATATTCCTATGATATGCTTTGGGTTGACGAGCTGCCCAATCCGTTAC 659  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 6167 STNSTSBYRANSTTFRNCYRASSTTTTCSRKYSCDCYNSRNCSTCCYCY 6108  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 660 AGGTAAATATCCTTTTGGCTCAGGTCGCTTAGCTTTACAGACGCGAGGTGTTGTTCC 719  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 6107 SSRSRNCSTTSSRCTBYRAYACSRSTSTDSRSRNCSTTCTBTCTTTTSTTTNST 6048  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 720 TAACATAGATAGATACATCAACGAGCCTTCTGGGTATTCCTGTAGTATTGGTGATGA 779  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 6047 CDNSTTBDRCDSRGRYSRRSCTRCSTCYTAKTBTBTYSYDAYDADCDRCYDASRC 5988  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 780 CAGGTGCTTGACCAACTATGCAACTGATTAGGAAAAGACTGTTTATCAATCCACTGTA 839  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 5987 STTYSYDAYSTTNTSRTBDBNSTNSDSTSRSTTTTADCCYTCNSRCYDTR 5928  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 840 ATGTATTAACAGATGTTCTGACAGATGCTACTTACTTGAACGAGCAAAACCGTGAA 899  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 5927 CTBYRATBNCSTNSTAKRNCSTSRCTTSCNSTSTNDCTBDRCYRARSRY 5868  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 900 CAAGTCTCTTTAGAGAGTCCATTTATTCGTTAAGAAATCAT--GNAACAATCCTTTT 957  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 5867 SNCRYSISTYDATTBTSRSGISNSTRCDYDAYSDYDASTDSTTSRSTTTTCTSTN 5808  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 958 GTAGCCCTATGACCACTAGTGTCTATGTTTATGATGCTTGTGTTTATCTGTGTTG 1017  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 5807 STSRDNRCTTTCYSSRYSTCYSTSYRAAKCYTYSRNSTNSTYDADNSTTB 5748  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1018 GATTTCTTTAGGCAATGTCAGAGATTTGATGGCTCAGGCTTTTACCTTTCTGCTG 1077  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 5747 STTTCYTTTTTSTNCYSTSYRANSTTBYRACYRACYCYTCTYSYDASTCYTSRG 5688  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1078 ATTATCTTCAATGTTGCTGTCGGAACATTCATTACATGCTTAAGCACCCTGCTCC 1137  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 5687 TBCVTTBYSTBNCRCNCYSSTYRANCYSSTYDABNSTYSTSTCTBYRAYSYDAYD 5628  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1138 TTCTTGTATCTCCGTTTATGGGTCGCAATTTGTTGCTCCTCAGACCCCTGAAATTA 1197  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 5627 ATBDCNSTRCTTTTTCNSTRYAKTNTNCTNCTTTCYRCSYCSYSTBT 5568  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 1198 TATTCCTTTTACTATT 1215  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 QY 5567 TTTTTCCTTTTTCCTTTT 5550  
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: October 2, 2003, 09:10:06  
 Job time : 385 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 09:01:58 ; Search time 81 Seconds  
(without alignments)  
6833.265 Million cell updates/sec

Title: US-09-469-200D-1

Perfect score: 1254

Sequence: 1 atggaacattataaaacct.....cacgtataaaattattataa 1254

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	604.6	48.2	1512	2	US-08-270-581-1
2	604.6	48.2	1512	4	US-09-146-893-1
3	61.6	4.9	2947	4	US-08-675-499A-1
4	53.8	4.3	536165	4	US-09-214-808-1
5	45.6	3.6	7218	1	US-08-232-463-14
6	37.8	3.0	2173	1	US-08-036-210-6
7	37.8	3.0	2173	2	US-08-449-609-6
8	37.8	3.0	2173	4	US-09-361-096A-6
9	37.8	3.0	2309	1	US-08-036-210-10
10	37.8	3.0	2309	2	US-08-449-609-10
11	37.8	3.0	2309	4	US-09-361-096A-10
12	37.8	3.0	2692	1	US-08-036-210-14
13	37.8	3.0	2692	2	US-08-449-609-14
14	37.8	3.0	2692	4	US-09-361-096A-14
15	37.8	3.0	3973	1	US-08-036-210-21
16	37.8	3.0	3973	2	US-08-449-609-21
17	37.8	3.0	3973	4	US-09-361-096A-21
18	37	3.0	1257	4	US-09-134-001C-1214
19	35.6	2.8	19124	4	US-08-487-826B-13
20	35.4	2.8	580073	4	US-08-545-528D-1
21	35	2.8	1371	4	US-09-328-352-3464
22	35	2.8	1664976	4	US-08-916-421B-1
23	34.6	2.8	1392	4	US-09-328-352-474
24	34.4	2.7	1830121	4	US-09-557-884-1
25	34.4	2.7	1830121	4	US-09-643-990A-1
26	34.2	2.7	1866	1	US-08-328-961-1
27	34.2	2.7	1866	1	US-08-462-397-1

c	28	34.2	2.7	2310	1	US-08-417-330A-15	Sequence 15, Appl
	29	34.2	2.7	2848	2	US-08-805-918-1	Sequence 1, Appl
	30	34.2	2.7	4106	2	US-08-702-572-14	Sequence 14, Appl
	31	34.2	2.7	4732	6	5521093-4	Patent No. 5521093
c	32	34.2	2.7	1664976	4	US-08-916-421B-1	Sequence 1, Appl
c	33	33.8	2.7	2169	4	US-09-434-408-3	Sequence 3, Appl
	34	33.6	2.7	2945	2	US-08-687-372-2	Sequence 2, Appl
	35	33.6	2.7	9060	3	US-08-378-313-20	Sequence 20, Appl
	36	33.4	2.7	2184	4	US-09-134-001C-1230	Sequence 1230, Appl
c	37	33.4	2.7	6755	3	US-08-931-999-4	Sequence 4, Appl
	38	33.2	2.6	2628	2	US-08-696-944-1	Sequence 1, Appl
c	39	33	2.6	2164	3	US-08-760-615-3	Sequence 1, Appl
c	40	33	2.6	11460	4	US-09-336-910A-1	Sequence 1, Appl
	41	32.8	2.6	65042	4	US-09-784-316-3	Sequence 3, Appl
	42	32.4	2.6	799	3	US-08-961-083-13	Sequence 13, Appl
	43	32.4	2.6	799	4	US-09-536-784-13	Sequence 13, Appl
	44	32.4	2.6	1242	4	US-09-328-352-101	Sequence 101, Appl
	45	32.4	2.6	1404	4	US-09-134-001C-2464	Sequence 2464, Appl

## ALIGNMENTS

RESULT 1  
US-08-270-581-1  
; Sequence 1, Application US/08270581  
; Patent No. 5856168  
; GENERAL INFORMATION:  
; APPLICANT: Weigel, Paul H.  
; APPLICANT: DeAngelis, Paul L.  
; APPLICANT: Papaconstantinou, John  
; TITLE OF INVENTION: Hyaluronate Synthase Gene and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/270,581  
; APPLICATION NUMBER: US/08/270,581  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: UTSG:161\PAR  
; TELEPHONE: 713-787-1400  
; TELEFAX: 713-789-2679  
; INFORMATION FOR SEQ ID NO: 1:  
; LENGTH: 1512 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 193..1449  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1488..1510  
; US-08-270-581-1

Query Match 48.2% Score 604.6; DB 2; Length 1512;

Best Local Similarity	68.7%;	Pred. No. 7.4e-169;
Matches	848;	Conservative
	0;	Mismatches
	384;	Indels
	3;	Gaps
QY	10	TTAAAAACCTCATACACTGTGTGGCTTTAGTATTTTGGGTACTGTTGATTATACGTC
Db	202	TTTAAAAAACTTTAATGTGTTTATCTCTTTATTTTGGATATCTATCTTGGATTATCTA
QY	70	AATGTTATCTCTTTGGTGCTTAAGAGNACGTTCACATTTATGGCTTTTGGCTGTAGCT
Db	262	AATATGTATCTATTTTGGAAAC---ATCAACTGATGAGAAATTTATGGAGTAAATTAATTAAC
QY	130	TACCTATTAGTCAAAATGTCCCTTATCTCTTTTACAAGCCATTTAAGGGAAGCGTGGG
Db	319	TATCTAGTTATCAAACTGGATATATCTTCCCTTTATGAGCCATTTAAGAGNAATCCAAAT
QY	190	CAATATATAGTGTCCACCAATTAATTCCTCTTATACGAGATGCTGTGATGCTGTAGAG
Db	379	GACTATAAGTGTCTGCTGTAATTCCTCTTATAATAGAGTGGCGATCAATTATTAGAA
QY	250	ACCTTAAAAAGTGTTCAGCAGCAAACTATCCCTTACGACAGAAATTTATGTTGTTGACGAT
Db	439	ACACITTAAGTGTGTAGCACACCTATCCGTTATCAGAATTTTATTTGTTGATGAT
QY	310	GGAGTGTGATGAGACAGGATTAATAGCGCATTAAGAGCATPATGTGCTGACATGCTGAC
Db	499	GGGAGTCAAAACACAGATGCAATCAATAATTAAGAGATGATGTAATAGAGAAGTGGAT
QY	370	CTATCAAGCAATGTCAATTTTCATCGGTACAGAGAAATCAAGGAACCGTCATGCCACAG
Db	559	ATTTCTGAAACGTTATCGTTCACCGTTCCTTGTCAATAAGGAAACGCCATGTCTCA
QY	430	GCGTGGGCGTTTGAAGATCAGACGCTGATGTCCTTTTGGACGTTGACTCAGATACTTAT
Db	619	GGGTGGGCAATTTGAAAGATCTGACGCTGACGTTTTTTTAAACGCTAGACTCAGATCTTAT
QY	490	ATCTACCTCATGCTTTAGAGAGTGTGTTAAAAACCTTTAATGACCACTGTTTGTGCT
Db	679	ATCTATCCAATGCCCTTAGAGAACTCCCTTAAAAAGCTTCAATGATGAGACAGTTTATGCT
QY	550	GGAGGGGTCACTTAAATGTCAGAAATACAGAAACCAATCTCTTAAACGCTTGACACAT
Db	739	GCAACAGGACATTTGAATGCTAGAAACACAGAACTAACTTATTAACGCCATTTACAGAT
QY	610	ATTCGCTATGATTAAGCTTTTGGGTTGAACAGCTGCCAATCCGTTTACAGGTAATATC
Db	799	ATCCGTTACGATTAATGCCCTTTGGGGTGGAGCGTGTCTCAATCATTAACAGGTAATAT
QY	670	CTTGTGTTGCTCAGGTCGCGTTAGCGTTTACAGCGCAGGTTGTTTCTTAACATAGAT
Db	859	TTAGTTTGCTCAGGACCATTCAGTATTTATCGAGCTGAGTGATTTCTTACTTCTAGAG
QY	730	AGATACATCAACAGACCTTCCCTGGGTATTCCTGTAAGTATGGTGAATGACAGGCTGTG
Db	919	CGCTATAAAATCAACAATCTCAGTGTGTACCTGTTTACCTGTTAGCATTTGGGATGATCGATGTTA
QY	790	ACCAACTATGCAACTGATTTAGGAAGAGCTGTTTATCAATCCACATGCTTAATGATATACA
Db	979	ACAAATATGCTATGATTTAGGACGACTGCTACCAATCACAGCTAGATGTGATCT
QY	850	GATGTTCTCTGACAAGATGCTACTTACTTTGAAGCAGCAAAACCGCTGGCAACAGTCCCTC
Db	1039	GATGTACCTTTCCAATTTAAAAGTTATTTTAAAGCACAAAATCGATGGAATAACTTTT
QY	910	TTTACAGAGTCCATTAATTCGTGTAGAAAAATCATGACAACTCCTTTTGTAGCCCTATGG
Db	1099	TTTAGAGAACTATTAATTCGTGTTTAAAAAAATCTTTCTTAATCCCACTGTTGCTTATGG
QY	970	ACCATACTTGAGTGTCTATCTTTATGATGCTGTGTTTATCTGTGTTGGATTTCTTTGTA
Db	1159	ACTATTTTCCAAGTCGTTATCTTTATGATGTTGATTCGCAATTTGGCAATCTTTGTTT
QY	1030	GCCATATCAGAGAAATTTGATGTGGCTCAGGGTTTTAGCCGTTTTCTGGTGAATATCTTCAT

Db 202 TTTAAAAAACTTTAAATGTTTATCCCTTTATTTTGTGATCTATCTGATTTATCTA 261  
QY 70 AATGTTTATCTCTTTGTTGCTAAAGAGAGCTTGCAATTTATGCTTTTTCGTGATAGT 129  
Db 262 AATATGATCTATTTTGAAC---ATCAACTGTAGGAATTTATGGAGTAATTAATAAC 318  
QY 130 TACCTATTAGTCAAAATGCTCTTATCTCTTTTATCAAGCCATTTAAGGAAGGCTGG 189  
Db 319 TATCTAGTATCAACTTGGATATCTCTTTATGAGCCATTTAAGGAATTCACAT 378  
QY 190 CAATATAGGTTGAGCCATTTATCCCTCTTTATACGAAGATGCTGAGTCATCTAGAG 249  
Db 379 GACTATAAAGTTGCTGTTGTAATCTCTTTATTAAGAGATGCGGAGTCATTTAGAA 438  
QY 250 ACCTTAAAGAGTTTCAGCAGCAACCTTATCCCTAGCAGAAATTTATGTTGAGAT 309  
Db 439 ACCTTAAAGAGTTTTCAGCAGCAACCTTATCCCTAGCAGAAATTTATGTTGATG 498  
QY 310 GGAAGTGTGATGAGCAGGATTTAAGCGCATTTGAAGACTATGTCGTGACACTGTGAC 369  
Db 499 GGGAGTCAACACAGATGCAATACAATTTAAGAGATGTAATAGAGAGTGGAT 558  
QY 370 CTATCAAGCAATGCTATGTCATCGGTGAGAAAATCAAGGAAGCGCTCATGCACAG 429  
Db 559 ATTGTGCAAGAGTTATGCTTACCGTTCCCTGTCATTAAGGAAGAAACGCCATGCTCA 618  
QY 430 GCGTGGGCTTTGAAGATCAGACGCTGATGCTTTTTCACCGTTGACTCAGATCTAT 489  
Db 619 GCGTGGGANTTGAAGATCTGACGCTGACGTTTTTAAACGCTAGACTCAGATCTAT 678  
QY 490 ATCTACCCCTGATGCTTTAGAGGATGTTTAAACCTTTAATGACCAACTGTTTTGCT 549  
Db 679 ATCTATCCAAATGCTTTAGAGAACTCTTAAGAGTCAATGATGAGACAGTTTATGCT 738  
QY 550 GCGAGGCTCAGCTTATGTCAGAAATGAGCAACCAATCTCTTAAACAGCTTTGACAGAT 609  
Db 739 GCAACAGGACATTTGAATGCTAGAAACAGCAAACTAATCTATTAAACGCACTACAGAT 798  
QY 610 ATTGCGTATGATATGCTTTTGGCTTGAAGAGCTGCCCAATCGGTTACAGTAATATC 669  
Db 799 ATCCGTACGATATGCTTTGGGTGAGGCTGCTCAATCAATTAACAGGATATAT 858  
QY 670 CTGTTTCTCAGTCCGCTTACGCTTTTACAGACGGAGGTTGTTTCTCAATACATAGAT 729  
Db 859 TTAGTTGCTCAGACCAATGATATTTATCGAGTGAAGTGAATTTCTCAATCTTAACTAGAG 918  
QY 730 AGATACATCAACAGACTCTCTGGGTATCTGTAAGTATGTTGTTGATGACAGGCTGTG 789  
Db 919 CGTATTAATAATCAACATCTCCTAGGTTTACCTGTAGCAATTTGGGATGATCGATTTA 978  
QY 790 ACCAATATGCACTGATTTAGAAAGACTGTTTATCAATCCACTGCTAAATGATTAACA 849  
Db 979 ACAATATATGATTTAGAGCGACTGTCTACCAATCAACAGCTAGATGATCT 1038  
QY 850 GATGTTCTGACAGATGCTACTTACTTGAGCAGCAAAACCGCTGGAACAGTCCCTC 909  
Db 1039 GATGACCTTTCCAAATTAAGAGTTTATTAAGCAACAAATTCGATGCAATAACTTTT 1098  
QY 910 TTTAGAGAGTCCATTTATCTTGTAAAGAAATCATGAACATCTTTTGTAGCCCTATGG 969  
Db 1099 TTTAGAGATCTATTTCTGTATAAANAATCTTCTTAATCCCATCGTTGCTTTATGG 1158  
QY 970 ACATATCTGAGGTGCTATGTTTATGATGCTTTTATCTGTTGGTGGATTTCTTTGTA 1029  
Db 1159 ACTATTTTCGAAGTCGTTATGTTTATGATGTTGATGTCGCAATTTGGGATCTTTTGT 1218  
QY 1030 GGCATGTCAGAGAAATTTGATGCTCAGGCTTTAGGCTTTTGTGGTGAATATCTTCAT 1089  
Db 1219 AATCAAGCTATTTCAATTTAGACCTTATTAACACTTTTGGCTTTTATCCATCATCTTATC 1278  
QY 1090 GTTGGCCCTGTGCGGAACATCATTAACATGCTTAAGCACCCTGCTCTCTCTGTTATCT 1149

Db 1279 GTTCTTTTATGCTGTAATGTTTCATTTATATATGTTGGTCAACATCTCTGCTAGTTTGTATCT 1338  
QY 1150 CCGTTTATGAGGTCGTCATTTGTTGCTCCTACAGCCCTTGAAATTAATATCTCTTTT 1209  
Db 1339 CCTCTGATGGAATATTAACACTTCTGTTGCTTACAGCCCTTAAACTTTATCTTTATGC 1398  
QY 1210 ACTATTAGAAATGCTGACTGGGGAACACGCTAAAAA 1244  
Db 1399 ACCATTAAAAATACGAATGGGAACACGCTAAAAA 1433  
RESULT 3  
US-08-675-499A-1  
; Sequence 1, Application US/08675499A  
; Patent No. 6492150  
; GENERAL INFORMATION:  
; APPLICANT: McDonald, J. A.  
; APPLICANT: Spicer, A. P.  
; APPLICANT: Augustin, M. L.  
; TITLE OF INVENTION: GENE ENCODING HYALURONAN  
; TITLE OF INVENTION: SYNTHASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwesman, Lundberg, Woessner & Kluth, P.A.  
; STREET: P.O. Box 2938  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/675,499A  
; FILING DATE: 03-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Woessner, Warren D  
; REGISTRATION NUMBER: 30,440  
; REFERENCE/DOCKET NUMBER: 150.170US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-359-3260  
; TELEFAX: 612-359-3263  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2947 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-675-499A-1  
Query Match 4.9%; Score 61.6; DB 4; Length 2947;  
Best Local Similarity 49.1%; Pred. No. 3,1e-08;  
Matches 222; Conservative 0; Mismatches 224; Indels 6; Gaps 2;  
QY 474 TGACTCAGATCTATATCTATCTACCTCATCTTACAGGAGCTTGTAAAACCTT---TAA 530  
Db 1140 TGACTCAGATCTATGCTTGACCTCGCTCATCTGTCGATGTTGGAGTGGTGAAGGCTTTAGAGA 1199  
QY 531 TGACCCAACTGTTTGTGTCGACGGGTCACTTAATGTCAGAAATAGACAAACCAATCT 590  
Db 1200 AGACCCCTATGTTGGAGGTGTTGGAGGAGATGCTCCACATTTTAAACAGATGATGCTTCG 1259  
QY 591 CTTAAACAGCTTGACAGATATGCTGCTATGATATGCTTTTGGCGTTGAACGAGCTGCCCA 650  
Db 1260 GATCTCTCTCCAGCAGCGTGAGATCTGGATGCTCTTTAATATAGAAAGGCGCTGCCA 1319

QY 651 ATCCGTACAGTAATATCTTTGCTCAGTCCGCTTAGCGTTTACAGACGCGAGT 710  
Db 1320 GCTTATTTTGGTGTGCTCAGTCAAGCGGTCTCTGGGAATGACAGAACTTCCTT 1379  
QY 711 GGTGTTCTTACATAGATAGATACATCAACAGACCTTCTGGGTATTTCTGTAAGTAT 770  
Db 1380 GCTGATGAATTTGGAAGACTGGTACATCAGGAATTCATGGGTAACCAATGCAGTTT 1439  
QY 771 TGGTATCAGAGTCTTGACCAACTATGCACTGATTTAGAAAGACTCTTTATCAATC 830  
Db 1440 TGGTACAGAGGACCTTACCACAGGGTGTGAGTCTGGGCTATCAACTAAATACAC 1499  
QY 831 CACTG---CTAATGATTATACAGATGTTCTGACAAAGATGCTACTTACTTGAAGCAGA 887  
Db 1500 GGTGCTGCTCAAGTCTTACTGAACTCCCATAGATAATCTAGATGGCTGACACAGA 1559  
QY 888 ARAACGCTGGACAAAGTCTTCTTTAGAGAGT 919  
Db 1560 GACCGATGGAGCAAGTCTTACTTCCGAGAGT 1591

RESULT 4  
US-09-214-808-1/C  
; Sequence 1, Application US/09214808A  
; Patent No. 6475793  
; GENERAL INFORMATION:  
; APPLICANT: Rosenthal, Andre  
; APPLICANT: Freiberg, Christoph  
; APPLICANT: Perret, Xavier Philippe  
; APPLICANT: Broughton, William John  
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 symbiotic  
; Patent No. 6475793  
; TITLE OF INVENTION: Plasmid  
; FILE REFERENCE: CARP0068  
; CURRENT APPLICATION NUMBER: US/09/214,808A  
; PRIOR FILING DATE: 1999-06-22  
; PRIORITY APPLICATION NUMBER: PCR/IB97/00950  
; PRIOR FILING DATE: 1997-07-10  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 536165  
; TYPE: DNA  
; ORGANISM: Rhizobium  
US-09-214-808-1

Query Match 4.3%; Score 53.8; DB 4; Length 536165;  
Best Local Similarity 46.1%; Pred. No. 7.8e-05;  
Matches 181; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 403 AAAATCAAGAAAGCGTCATGCACAGCGCTGGGCTTTGAAAGATCAGACGCTGATGC 462  
Db 156991 AAGAACGTCGGAAGCGCAAGCGCAGATTGTGCGATACGGAATCATCGGAGATTG 156932  
QY 463 TTTTTCACGCTGACCTACAGATCTATATCTACCTGATGCTTTAGAGGAGTTGTAATA 522  
Db 156931 GTGCTACAGCTTGACTCGGACAGACCACTTGGCCGAGCAGTAGTCAGAACTTCCCTG 156872  
QY 523 ACCCTTATGACCACTGCTTTTGTGTCGAGCGGTTCACCTTATGATGCAAGTAAGACAA 582  
Db 156871 AAGATGTACAGTCCCGGTCGGCGCGGATGGGTGAGTGTAGCGGCGAGCACCGCAGC 156812  
QY 583 ACCAATCTTTACAGCTTGACAGATATTCGCTATGATATGCTTTTGGCGTTGACACGA 642  
Db 156811 GACATATGCTGACGCGGTGATGACATGAGTACTGGCTCGCTGCAACAGAGACGC 156752  
QY 643 GCTGCCAATCCCTTACAGTAATATCTTTGTCAGTCGCTGCTTACGCTTTACAGA 702  
Db 156751 GCAGCAGAGCTCGCTTTTGAGCGGTTATGTTGTCGCGCCGCTGCTGCTGATGACGG 156692  
QY 703 CCGAGGTGCTGTTCTTCAACATAGATAGATACATCAACAGACCTTCTTGGGATTCCT 762

Db 156691 CGGTCCGCACTCTCTATGCTGCTGTAATAATACAGAGCACTGTTTCGAGGAGGCCA 156632  
QY 763 GTAAATGTTGGTATGATGACAGGTGCTTGAACCAAC 795  
Db 156631 AGCGACTTCGGGGAAGACCGCCACCTTCAATC 156599

RESULT 5  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZ9pt-Fls  
US-08-232-463-14

Query Match 3.6%; Score 45.6; DB 1; Length 7218;  
Best Local Similarity 3.4%; Pred. No. 0.0026;  
Matches 12; Conservative 197; Mismatches 141; Indels 0; Gaps 0;

QY 904 TCCTCTTTAGAGAGTCCATATTTCTGTAAGAAATCATGAACAATCCTTTGTAGCC 963  
Db 1101 YY 1160  
QY 964 CTATGACCACTTATGAGTCTATGTTTANGATCTGTTTATCTGTGPGGATTC 1023  
Db 1161 YY 1220  
QY 1024 TTGTAGCAATGTCAGAGAAATTCATTTGGCTCAGGTTTAGCCTTCTGCTGATATC 1083  
Db 1221 YY 1280  
QY 1084 TCCATTTGCTGCTGCGGAACATTCATTACATGCTTAAGCACCGCTGCTCTTG 1143

Db 1281 YY 1340  
 QY 1144 TTATCCCGCTTTTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1203  
 Db 1341 YY 1400  
 QY 1204 CTTTCTATTAGAAATGCTGACTGGGACAGCTAAATAATTATTATA 1253  
 Db 1401 YY 1450

RESULT 6  
 US-08-036-210-6  
 ; Sequence 6, Application US/08036210  
 ; Patent No. 595233  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Moller, Niels P.H.  
 ; APPLICANT: Moller, Karin B.  
 ; APPLICANT: Ullrich, Axel  
 ; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
 ; NUMBER OF SEQUENCES: 45  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/036,210  
 ; FILING DATE: 23-MAR-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mirock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; REFERENCE/DOCKET NUMBER: 7683-025  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-790-9090  
 ; TELEFAX: 212-869-8864/9741  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2173 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 1630  
 ; OTHER INFORMATION: /note= "N=x-unknown nucleotide"  
 US-08-036-210-6

Query Match 3.08; Score 37.8; DB 1; Length 2173;  
 Best Local Similarity 48.4%; Pred. No. 0.29;  
 Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;  
 QY 874 TACTTGAAGCAGCAAAACCGCTGGAAACAGTCCTCTTTAGAGAGTCATTTTCGTT 933  
 Db 1877 TACTTGGTGGACTGAAACAAAGAAATTAACCATCTATAGCAAAATTCAGGTTCTTTA 1936  
 QY 934 AAGAAATCATGAACAATCCTTTTGTAGCCCTATGGACCATCTTCAGGTCTCTATGTT 993  
 Db 1937 TAGAAAATCTTTCAGCCCTCCATCTTATTAAATAGTGAATGTAAGTTTGAATAT 1996  
 QY 994 ATGATGCTGTTTATTCTGTGGTGGAATTTCTTTGTAGGCAATGTCAGAGAAATTGATTGG 1053

Db 1997 ATGAACCTCATTTTGTCTAGATATTTCAATTAAAGATAAATAAGTATTAATATCTCT 2056  
 QY 1054 CTCAGGGTTTACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1090  
 Db 2057 CTATGATAAGAAATATATCTTATGCTTATCTTATTCGCTG 2093  
 RESULT 7  
 US-08-449-609-6  
 ; Sequence 6, Application US/08449609  
 ; Patent No. 5952212  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Moller, Niels P.H.  
 ; APPLICANT: Moller, Karin B.  
 ; APPLICANT: Ullrich, Axel  
 ; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
 ; NUMBER OF SEQUENCES: 45  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/449,609  
 ; FILING DATE: 24-MAY-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/036,210  
 ; FILING DATE: 23-MAR-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mirock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; REFERENCE/DOCKET NUMBER: 7683-025  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-790-9090  
 ; TELEFAX: 212-869-8864/9741  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2173 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: 1630  
 ; OTHER INFORMATION: /note= "N=x-unknown nucleotide"  
 US-08-449-609-6

Query Match 3.0%; Score 37.8; DB 2; Length 2173;  
 Best Local Similarity 48.4%; Pred. No. 0.29;  
 Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;  
 QY 874 TACTTGAAGCAGCAAAACCGCTGGAAACAGTCCTCTTTAGAGAGTCATTTTCGTT 933  
 Db 1877 TACTTGGTGGACTGAAACAAAGAAATTAACCATCTATAGCAAAATTCAGGTTCTTTA 1936  
 QY 934 AAGAAATCATGAACAATCCTTTTGTAGCCCTATGGACCATCTTCAGGTCTCTATGTT 993  
 Db 1937 TAGAAAATCTTTCAGCCCTCCATCTTATTAAATAGTGAATGTAAGTTTGAATAT 1996  
 QY 994 ATGATGCTGTTTATTCTGTGGTGGAATTTCTTTGTAGGCAATGTCAGAGAAATTGATTGG 1053

Db 1997 ATGAACATCTTTTCATAGATTTCAATTAAGAGTAATAAATAGATTAATTAATGCTCTT 2056  
QY 1054 CTCAGGTTTACGCTTCTGTCGATTAATTCATG 1090  
Db 2057 CTATGATAAGAGTATATCTTATGCTTATTCGCTG 2093

## RESULT 8

US-09-361-096A-6  
; Sequence 6, Application US/09361096A  
; Patent No. 6492495  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE  
; FILE REFERENCE: 038602/0686  
; CURRENT APPLICATION NUMBER: US/09/361.096A  
; CURRENT FILING DATE: 1999-07-26  
; PRIOR APPLICATION NUMBER: 08/449,609  
; PRIOR FILING DATE: 1995-05-24  
; PRIOR APPLICATION NUMBER: 08/036,210  
; PRIOR FILING DATE: 1995-03-23  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2173  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PTP-S31C  
; NAME/KEY: modified base  
; LOCATION: (1)..(1630)  
; OTHER INFORMATION: n = unknown nucleotide  
; NAME/KEY: CDS  
; LOCATION: (65)..(1066)  
US-09-361-096A-6

Query Match 3.0%; Score 37.8; DB 4; Length 2173;  
Best Local Similarity 48.4%; Pred. No. 0.29;  
Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;  
QY 874 TACTGAGCAGCAAAACCGCTGGAAAGTCTCTTTAGAGAGTCCATTAATTTCTGTT 933  
Db 1877 TACTGAGTGAAGTGAACCAAGAAATTAACCATCTATAGCAATTCAGGTTTCTTA 1936  
QY 934 AAGAAATCATGAAACATCTTTTGTAGCCCTATGGACCATACTTGGTGCTATGTTT 993  
Db 1937 TGAATAATCTTTCAGCCTCCATCTTATTAATAGTAGTGAATGTTTGAATAT 1996  
QY 994 ATGATGCTTTTATCTGTGCTGCTGATTTCTTTGTAGCAATGTCAGAGAAATTTGTTG 1053  
Db 1997 ATGAACATCTTTTGTCATAGATTTCAATTAAGAGTAATAAATAGTATTAATTAATGCTCTT 2056  
QY 1054 CTCAGGTTTACGCTTCTGTCGATTAATTCATG 1090  
Db 2057 CTATGATAAGAGTATATCTTATGCTTATTCGCTG 2093

## RESULT 9

US-08-036-210-10  
; Sequence 10, Application US/08036210  
; Patent No. 5585233  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas

CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/036,210  
FILING DATE: 23-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2309 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1766  
OTHER INFORMATION: /note= "N=x-unknown nucleotide"  
US-08-036-210-10

Query Match 3.0%; Score 37.8; DB 1; Length 2309;  
Best Local Similarity 48.4%; Pred. No. 0.3;  
Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;  
QY 874 TACTGAGCAGCAAAACCGCTGGAAAGTCTCTTTAGAGAGTCCATTAATTTCTGTT 933  
Db 2013 TACTGAGTGAAGTGAACCAAGAAATTAACCATCTATAGCAATTCAGGTTTCTTA 2072  
QY 934 AAGAAATCATGAAACATCTTTTGTAGCCCTATGGACCATACTTGGTGCTATGTTT 993  
Db 2073 TGAATAATCTTTCAGCCTCCATCTTATTAATAGTAGTGAATGTTTGAATAT 2132  
QY 994 ATGATGCTTTTATCTGTGCTGCTGATTTCTTTGTAGCAATGTCAGAGAAATTTGTTG 1053  
Db 2133 ATGAACATCTTTTGTCTCATAGATTTCAATTAAGAGTAATAAATAGTATTAATTAATGCTCTT 2192  
QY 1054 CTCAGGTTTACGCTTCTGTCGATTAATTCATG 1090  
Db 2193 CTATGATAAGAGTATATCTTATGCTTATTCGCTG 2229

## RESULT 10

US-08-449-609-10  
; Sequence 10, Application US/08449609  
; Patent No. 5952212  
; GENERAL INFORMATION:  
; APPLICANT: MOLLER, NIELS P.H.  
; APPLICANT: MOLLER, KARIN B.  
; APPLICANT: ULLRICH, AXEL  
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2309 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1766
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"
;
US-08-449-609-10

Query Match 3.0%; Score 37.8; DB 2; Length 2309;
Best Local Similarity 48.4%; Pred. No. 0.3;
Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 874 TACTTGAAGCAGCAAAACCGCTGGAAACGCTCTTTTGAAGAGTCATTAATTTCTGTT 933
Db 2013 TACTTGGTGAACCTGAAACAAAGAAATTAACCATCTATAGCAAAATTCAGGTTCTTTA 2072

QY 934 AAGAAATCATGAACAATCCCTTTTGTAGCCCTTAGGACCATACTTGGAGTGTCTATGTT 993
Db 2073 TAGAAATCTTTTCAGCTCCATCTTATTAATAGTACAAATGTGTAAGTTTGAATTAT 2132

QY 994 ATGATGCTGTTTATCTGTTGGTGAATTTCTTTAGGCAATGTCAGAGAAATTTGATTGG 1053
Db 2133 ATGACTATTTTGTATAGATTTCAATTAGAGTAATAAATAGTATTAATTAATGCTCTT 2192

QY 1054 CTCAGGGTTTAGCCTTTCTGGTGATTAATCTTCATTG 1090
Db 2193 CTATGATAAGAAGTATATCTTATGCTTATTTCCGCTG 2229

RESULT 11
US-09-361-096A-10
; Sequence 10, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/09/361,096A
; CURRENT FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2309 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1766
; OTHER INFORMATION: /note= "N=x-unknown nucleotide"
;
US-08-449-609-10

Query Match 3.0%; Score 37.8; DB 4; Length 2309;
Best Local Similarity 48.4%; Pred. No. 0.3;
Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 874 TACTTGAAGCAGCAAAACCGCTGGAAACGCTCTTTTGAAGAGTCATTAATTTCTGTT 933
Db 2013 TACTTGGTGAACCTGAAACAAAGAAATTAACCATCTATAGCAAAATTCAGGTTCTTTA 2072

QY 934 AAGAAATCATGAACAATCCCTTTTGTAGCCCTTAGGACCATACTTGGAGTGTCTATGTT 993
Db 2073 TAGAAATCTTTTCAGCTCCATCTTATTAATAGTACAAATGTGTAAGTTTGAATTAT 2132

QY 994 ATGATGCTGTTTATCTGTTGGTGAATTTCTTTAGGCAATGTCAGAGAAATTTGATTGG 1053
Db 2133 ATGACTATTTTGTATAGATTTCAATTAGAGTAATAAATAGTATTAATTAATGCTCTT 2192

QY 1054 CTCAGGGTTTAGCCTTTCTGGTGATTAATCTTCATTG 1090
Db 2193 CTATGATAAGAAGTATATCTTATGCTTATTTCCGCTG 2229

RESULT 12
US-08-036-210-14
; Sequence 14, Application US/08036210
; Patent No. 5585233
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/08/036,210
; CURRENT FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,210
; FILING DATE: 23-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2692 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

```
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2149
; OTHER INFORMATION: /note= "N-x-unknown nucleotide"
US-08-036-210-14

Query Match          3.0%; Score 37.8; DB 1; Length 2692;
Best Local Similarity 48.4%; Pred. No. 0.32;
Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 874 TACTTGAAGCAGCAAAACCGCTGGAACAGTCTCTTTAGAGAGTCCATTATTTCTGTT 933
Db 2396 TACTTGTGGAACTGAAACAAGAAATTAACCATCTATAGCAAAATCAAGGTTTCTTTA 2455

QY 934 AAGAAATCATGAACAATCCTTTTGTAGCCCTATGGACCATCTTGGAGTGTCTATGTTT 993
Db 2456 TAGAAATCTTTTCAGCCTCCATCTTATTAATAGTGACAATGTGTAAGTTTGAATTAT 2515

QY 994 ATGATGCTTGTATTCTGTGGTGGATTCTTTGTAGGCAATGTCAGAGAATTTGATGG 1053
Db 2516 ATGAATCATTTTGTCTATAGATTTCATTAAGAGTAATAAATAGTATTATATGCTCTT 2575

QY 1054 CTCAGGTTTTCAGCCTTCTGTGATTAATCTTCATTG 1090
Db 2576 CTATGATAAGAGTATATCTATGCTTATTTCCGCTG 2612

RESULT 13
US-08-449-609-14
; Sequence 14, Application US/08449609
; Patent No. 5952212
; GENERAL INFORMATION:
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,609
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,210
; FILING DATE: 23-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2692 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown

; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2149
; OTHER INFORMATION: /note= "N-x-unknown nucleotide"
US-08-449-609-14

Query Match          3.0%; Score 37.8; DB 2; Length 2692;
Best Local Similarity 48.4%; Pred. No. 0.32;
Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 874 TACTTGAAGCAGCAAAACCGCTGGAACAGTCTCTTTAGAGAGTCCATTATTTCTGTT 933
Db 2396 TACTTGTGGAACTGAAACAAGAAATTAACCATCTATAGCAAAATCAAGGTTTCTTTA 2455

QY 934 AAGAAATCATGAACAATCCTTTTGTAGCCCTATGGACCATCTTGGAGTGTCTATGTTT 993
Db 2456 TAGAAATCTTTTCAGCCTCCATCTTATTAATAGTGACAATGTGTAAGTTTGAATTAT 2515

QY 994 ATGATGCTTGTATTCTGTGGTGGATTCTTTGTAGGCAATGTCAGAGAATTTGATGG 1053
Db 2516 ATGAATCATTTTGTCTATAGATTTCATTAAGAGTAATAAATAGTATTATATGCTCTT 2575

QY 1054 CTCAGGTTTTCAGCCTTCTGTGATTAATCTTCATTG 1090
Db 2576 CTATGATAAGAGTATATCTATGCTTATTTCCGCTG 2612

RESULT 14
US-09-361-096A-14
; Sequence 14, Application US/09361096A
; Patent No. 6492495
; GENERAL INFORMATION:
; APPLICANT: MOLLER, NIELS P.H.
; APPLICANT: MOLLER, KARIN B.
; APPLICANT: ULLRICH, AXEL
; TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE PHOSPHATASE
; FILE REFERENCE: 038602/0686
; CURRENT APPLICATION NUMBER: US/09/361,096A
; CURRENT FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 08/449,609
; PRIOR FILING DATE: 1995-05-24
; PRIOR APPLICATION NUMBER: 08/036,210
; PRIOR FILING DATE: 1995-03-23
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 14
; LENGTH: 2692
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1)..(2692)
; OTHER INFORMATION: n = unknown nucleotide
; NAME/KEY: CDS
; LOCATION: (92..139, 259..1414)
US-09-361-096A-14

Query Match          3.0%; Score 37.8; DB 4; Length 2692;
Best Local Similarity 48.4%; Pred. No. 0.32;
Matches 105; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 874 TACTTGAAGCAGCAAAACCGCTGGAACAGTCTCTTTAGAGAGTCCATTATTTCTGTT 933
Db 2396 TACTTGTGGAACTGAAACAAGAAATTAACCATCTATAGCAAAATCAAGGTTTCTTTA 2455

QY 934 AAGAAATCATGAACAATCCTTTTGTAGCCCTATGGACCATCTTGGAGTGTCTATGTTT 993
Db 2456 TAGAAATCTTTTCAGCCTCCATCTTATTAATAGTGACAATGTGTAAGTTTGAATTAT 2515

QY 994 ATGATGCTTGTATTCTGTGGTGGATTCTTTGTAGGCAATGTCAGAGAATTTGATGG 1053
Db 2516 ATGAATCATTTTGTCTATAGATTTCATTAAGAGTAATAAATAGTATTATATGCTCTT 2575
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 09:10:12 ; Search time 345 Seconds  
(without alignments)  
9271.834 Million cell updates/sec

Title: US-09-469-200D-1  
Perfect score: 1254  
Sequence: 1 atgagacattataaacct.....caogtaaaattattataa 1254

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications.NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	1254	100.0	1254	10	US-09-879-959-1	Sequence 1, Appli
2	1254	100.0	1254	14	US-10-011-768B-1	Sequence 1, Appli
3	1254	100.0	1254	14	US-10-011-771B-1	Sequence 1, Appli
4	1254	100.0	1254	14	US-10-172-527-1	Sequence 1, Appli
5	1225.4	97.7	1251	12	US-10-326-185-1	Sequence 1, Appli
6	1152	91.9	5158	12	US-10-326-185-108	Sequence 108, App
7	623.2	49.7	1251	12	US-10-326-185-102	Sequence 102, App
8	623.2	49.7	3466	14	US-10-172-527-11	Sequence 11, Appl
9	604.6	48.2	1440	14	US-10-172-527-13	Sequence 13, Appl
10	604.6	48.2	1512	14	US-10-124-222-1	Sequence 1, Appli
11	604.6	48.2	1512	14	US-10-117-795-1	Sequence 1, Appli
12	601.4	48.0	1257	12	US-10-326-185-92	Sequence 92, Appl
13	282.8	22.6	1200	14	US-10-172-527-17	Sequence 17, Appl
14	61.6	4.9	1659	11	US-09-902-939-2	Sequence 2, Appli
15	61.6	4.9	4134	14	US-10-262-526-3	Sequence 3, Appli
16	54	4.3	54	12	US-10-326-185-39	Sequence 39, Appl

c	17	53.8	4.3	536165	11	US-09-939-964-1	Sequence 1, Appli
	18	49.4	3.9	1665	11	US-09-902-939-3	Sequence 3, Appli
	19	47.2	3.8	3003	11	US-09-918-624B-64	Sequence 64, Appli
	20	47.2	3.8	3003	14	US-10-262-526-1	Sequence 1, Appli
	21	45.6	3.6	4018	12	US-09-814-353-19942	Sequence 19942, A
	22	43	3.4	1752	11	US-09-902-939-1	Sequence 1, Appli
	23	43	3.4	3387	14	US-10-084-817-184	Sequence 184, App
	24	43	3.4	7488	12	US-10-311-455-1085	Sequence 1085, Ap
	25	43	3.4	9899	12	US-10-311-455-2416	Sequence 2416, Ap
	26	41.6	3.3	1740	10	US-09-879-959-8	Sequence 8, Appli
	27	41.6	3.3	1740	14	US-10-011-768B-7	Sequence 7, Appli
	28	41.6	3.3	1740	14	US-10-011-771B-7	Sequence 7, Appli
	29	41.6	3.3	1740	14	US-10-172-527-8	Sequence 8, Appli
	30	41	3.3	6301	12	US-10-311-455-26	Sequence 26, Appli
	31	39	3.1	3528	10	US-09-070-927A-189	Sequence 189, App
	32	38.8	3.1	888	12	US-10-032-585-6433	Sequence 6433, Ap
	33	38.8	3.1	9817	12	US-10-311-455-1341	Sequence 1341, Ap
	34	38	3.0	6485	12	US-10-311-455-1781	Sequence 1781, Ap
	35	38	3.0	7049	12	US-10-311-455-130	Sequence 130, App
	36	38	3.0	7049	12	US-10-240-452-6	Sequence 6, Appli
	37	38	3.0	7119	12	US-10-311-455-266	Sequence 266, App
	38	38	3.0	15306	12	US-10-240-453-184	Sequence 184, App
	39	38	3.0	15306	14	US-10-239-676-162	Sequence 162, App
	40	37.8	3.0	2173	12	US-10-314-232-6	Sequence 6, Appli
	41	37.8	3.0	2309	12	US-10-314-232-10	Sequence 10, Appli
	42	37.8	3.0	2692	12	US-10-314-232-14	Sequence 14, Appli
	43	37.8	3.0	3973	12	US-10-314-232-21	Sequence 21, Appli
	44	37.6	3.0	9243	12	US-10-240-453-127	Sequence 127, App
	45	37.4	3.0	600	13	US-10-027-632-84036	Sequence 84036, A

#### ALIGNMENTS

#### RESULT 1

US-09-879-959-1  
; Sequence 1, Application US/09879959  
; Patent No. US20020160489A1  
; GENERAL INFORMATION:  
; APPLICANT: Weigel, Paul H  
; APPLICANT: Kumari, Kahana  
; APPLICANT: DeAngelis, Paul  
; TITLE OF INVENTION: STREPTOCOCCUS EQUISIMILIS HYALURONAN SYNTHASE GENE AND EXPRESS  
; FILE REFERENCE: 3554 049  
; CURRENT APPLICATION NUMBER: US/09/879,959  
; PRIOR FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: 09/469,200  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 09/178,851  
; PRIOR FILING DATE: 1998-10-26  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 1  
; LENGTH: 1254  
; TYPE: DNA  
; ORGANISM: Streptococcus equisimilis  
US-09-879-959-1

Query Match 100.0%; Score 1254; DB 10; Length 1254;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAGACATTAAAAACCTCATACCTCTGCGCCCTTAGTATTTTGGTACGTTTG 60  
|||||  
Db 1 ATGAGACATTAAAAACCTCATACCTCTGCGCCCTTAGTATTTTGGTACGTTTG 60  
|||||  
QY 61 ATTACGTCATATTTATCTCTTTGGTGTAAAGAGCTTGTCAATTATGCGCTTTTG 120  
|||||  
Db 61 ATTACGTCATATTTATCTCTTTGGTGTAAAGAGCTTGTCAATTATGCGCTTTTG 120  
|||||  
QY 121 CTGACGCTTACCTATTAGTCAAAATGCTTCCTATCTCTTTTACAAGCCATTAGGA 180  
|||||

```
Db 121 CTGATAGCTACCTATTAGTCAAAATGTCCTTATCCCTTTTACAGCCATTAAAGGA 180
QY 181 AGGCTGGCAATATAAGTTGACGCCATTATCCCTCTTATAACGAAGATGCTGAGTCA 240
| 181 AGGCTGGCAATATAAGTTGACGCCATTATCCCTCTTATAACGAAGATGCTGAGTCA 240
QY 241 TTGCTAGAGACCTTAAAGTGTTCAGCAGCAACCTATCCCTAGCAGAAATTTATGTT 300
Db 241 TTGCTAGAGACCTTAAAGTGTTCAGCAGCAACCTATCCCTAGCAGAAATTTATGTT 300
QY 301 GTTGACGATGGAAGTGTGATGAGACAGTATTAAAGCATTAAGCGATTGAAGACTATGTCGTGAC 360
Db 301 GTTGACGATGGAAGTGTGATGAGACAGTATTAAAGCATTAAGCGATTGAAGACTATGTCGTGAC 360
QY 361 ACTGGTGACCTATCAAGCAATGTCATTGTCATCGGTGAGAGAAAATCAAGGAAGCGT 420
Db 361 ACTGGTGACCTATCAAGCAATGTCATTGTCATCGGTGAGAGAAAATCAAGGAAGCGT 420
QY 421 CATGCACAGCGCTGGCCCTTTGAAGATCAGACGCTGATGCTTTTGAACGTTGACTCA 480
Db 421 CATGCACAGCGCTGGCCCTTTGAAGATCAGACGCTGATGCTTTTGAACGTTGACTCA 480
QY 481 GATACCTATATCAACCTGATGCTTTAGAGGAGTGTAAACACCTTTAATGACCCAACT 540
Db 481 GATACCTATATCAACCTGATGCTTTAGAGGAGTGTAAACACCTTTAATGACCCAACT 540
QY 541 GTTTTGTGCGGACGGTCAACCTTAATGTCAGAAATAGACAAACCACTCTTAACACGC 600
Db 541 GTTTTGTGCGGACGGTCAACCTTAATGTCAGAAATAGACAAACCACTCTTAACACGC 600
QY 601 TTGACAGATATTCGCTATGATAATGCTTTTGGCGTTGAGGAGCTGCCCAATCGGTTACA 660
Db 601 TTGACAGATATTCGCTATGATAATGCTTTTGGCGTTGAGGAGCTGCCCAATCGGTTACA 660
QY 661 GGTAAATCTGTTGTCAGGCTGAGGCTGAGGTTTACAGACGCGAGTGCTGTCCT 720
Db 661 GGTAAATCTGTTGTCAGGCTGAGGCTGAGGTTTACAGACGCGAGTGCTGTCCT 720
QY 721 AACATAGATAGATACATCAACGACGCTTCTGGGTATTCGTAGTATGTTGATGAC 780
Db 721 AACATAGATAGATACATCAACGACGCTTCTGGGTATTCGTAGTATGTTGATGAC 780
QY 781 AGGTGCTTGACCAACTATGCAACTGATTTAGGAAGAGCTGTTATCAATCCACTGCTAAA 840
Db 781 AGGTGCTTGACCAACTATGCAACTGATTTAGGAAGAGCTGTTATCAATCCACTGCTAAA 840
QY 841 TGTATTACAGATGCTTGGACAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAAC 900
Db 841 TGTATTACAGATGCTTGGACAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAAC 900
QY 901 AAGTCTCTTTTAGAGAGTCCATTATTTCTGTTAAGAAAATCATGAACAATCCTTTTGT 960
Db 901 AAGTCTCTTTTAGAGAGTCCATTATTTCTGTTAAGAAAATCATGAACAATCCTTTTGT 960
QY 961 GGCCTATGGACCAACTGAGGCTCTATGTTTATGATGCTGTTTATCTGCTGGTGGAT 1020
Db 961 GGCCTATGGACCAACTGAGGCTCTATGTTTATGATGCTGTTTATCTGCTGGTGGAT 1020
QY 1021 TTTCTTTGAGCAATGTCAGAGAAATTTGATTGGCTCAGGTTTTTACGTTTCTGGTGATT 1080
Db 1021 TTTCTTTGAGCAATGTCAGAGAAATTTGATTGGCTCAGGTTTTTACGTTTCTGGTGATT 1080
QY 1081 ATCTTCAATGTTGCCCTGTGCGGAACATTCATTACATGCTTTAAGCACCGCTGTCCTTC 1140
Db 1081 ATCTTCAATGTTGCCCTGTGCGGAACATTCATTACATGCTTTAAGCACCGCTGTCCTTC 1140
QY 1141 TTGTTATCTCGGTTTATGGGGTGTGCATTTGTTTGTCTTACAGCCCTTGAATATAT 1200
Db 1141 TTGTTATCTCGGTTTATGGGGTGTGCATTTGTTTGTCTTACAGCCCTTGAATATAT 1200
QY 1201 TCTCTTTTACTATTAGAAATGCTGACTGGGGAACACGCTAAAATAATTTATATA 1254
Db 1201 TCTCTTTTACTATTAGAAATGCTGACTGGGGAACACGCTAAAATAATTTATATA 1254
```

## RESULT 2

```
US-10-011-768B-1
; Sequence 1, Application US/10011768B
; Publication No. US20030073221A1
; GENERAL INFORMATION:
; APPLICANT: Weigel, Paul H.
; APPLICANT: DeAngelis, Paul
; APPLICANT: Kumari, Kshama
; TITLE OF INVENTION: Hyaluronan Synthase Gene and Uses Thereof
; FILE REFERENCE: 3554.011
; CURRENT APPLICATION NUMBER: US/10/011,768B
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: US 09/178,851
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: US 60/064,435
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Streptococcus equisimilis
US-10-011-768B-1

Query Match 100.0%; Score 1254; DB 14; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGAACATTAATAAACCTCATACCTGTTGGCGCTTAAGAGAGCTTGTCAATTTATGGCTTTTG 120
| 1 ATGAGAACATTAATAAACCTCATACCTGTTGGCGCTTAAGAGAGCTTGTCAATTTATGGCTTTTG 60
Db 1 ATGAGAACATTAATAAACCTCATACCTGTTGGCGCTTAAGAGAGCTTGTCAATTTATGGCTTTTG 60

QY 61 ATTTAGCTCAATGTTATCTCTTTGGTGCTTAAGAGAGCTTGTCAATTTATGGCTTTTG 120
| 61 ATTTAGCTCAATGTTATCTCTTTGGTGCTTAAGAGAGCTTGTCAATTTATGGCTTTTG 120
Db 61 ATTTAGCTCAATGTTATCTCTTTGGTGCTTAAGAGAGCTTGTCAATTTATGGCTTTTG 120

QY 121 CTGATAGCTTACCTTATAGTCAAAATGTCCTTATCCCTTTTACAGCCATTAAAGGA 180
| 121 CTGATAGCTTACCTTATAGTCAAAATGTCCTTATCCCTTTTACAGCCATTAAAGGA 180
Db 121 CTGATAGCTTACCTTATAGTCAAAATGTCCTTATCCCTTTTACAGCCATTAAAGGA 180

QY 181 AGGCTGGGCAATATAAGTTGACGCCATTATCCCTCTTATAACGAAGATGCTGAGTCA 240
| 181 AGGCTGGGCAATATAAGTTGACGCCATTATCCCTCTTATAACGAAGATGCTGAGTCA 240
Db 181 AGGCTGGGCAATATAAGTTGACGCCATTATCCCTCTTATAACGAAGATGCTGAGTCA 240

QY 241 TTGCTAGACACCTTAAAGTGTTCAGCAGCAACCTATCCCTAGCAGAAATTTATGTT 300
| 241 TTGCTAGACACCTTAAAGTGTTCAGCAGCAACCTATCCCTAGCAGAAATTTATGTT 300
Db 241 TTGCTAGACACCTTAAAGTGTTCAGCAGCAACCTATCCCTAGCAGAAATTTATGTT 300

QY 301 GTTGACGATGGAAGTGTGATGAGACAGTATTAAAGCATTAAGCGATTGAAGACTATGTCGTGAC 360
| 301 GTTGACGATGGAAGTGTGATGAGACAGTATTAAAGCATTAAGCGATTGAAGACTATGTCGTGAC 360
Db 301 GTTGACGATGGAAGTGTGATGAGACAGTATTAAAGCATTAAGCGATTGAAGACTATGTCGTGAC 360

QY 361 ACTGGTGACCTATCAAGCAATGTCATTGTCATCGGTGAGAGAAAATCAAGGAAGCGT 420
| 361 ACTGGTGACCTATCAAGCAATGTCATTGTCATCGGTGAGAGAAAATCAAGGAAGCGT 420
Db 361 ACTGGTGACCTATCAAGCAATGTCATTGTCATCGGTGAGAGAAAATCAAGGAAGCGT 420

QY 421 CATGCACAGCGCTGGCCCTTTGAAGATCAGACGCTGATGCTTTTGAACGTTGACTCA 480
| 421 CATGCACAGCGCTGGCCCTTTGAAGATCAGACGCTGATGCTTTTGAACGTTGACTCA 480
Db 421 CATGCACAGCGCTGGCCCTTTGAAGATCAGACGCTGATGCTTTTGAACGTTGACTCA 480

QY 481 GATACCTATATCAACCTGATGCTTTAGAGGAGTGTAAACACCTTTAATGACCCAACT 540
| 481 GATACCTATATCAACCTGATGCTTTAGAGGAGTGTAAACACCTTTAATGACCCAACT 540
Db 481 GATACCTATATCAACCTGATGCTTTAGAGGAGTGTAAACACCTTTAATGACCCAACT 540

QY 541 GTTTTGTGCGGACGGTCAACCTTAATGTCAGAAATAGACAAACCACTCTTAACACGC 600
| 541 GTTTTGTGCGGACGGTCAACCTTAATGTCAGAAATAGACAAACCACTCTTAACACGC 600
Db 541 GTTTTGTGCGGACGGTCAACCTTAATGTCAGAAATAGACAAACCACTCTTAACACGC 600

QY 601 TTGACAGATATTCGCTATGATAATGCTTTTGGCGTTGAGGAGCTGCCCAATCGGTTACA 660
| 601 TTGACAGATATTCGCTATGATAATGCTTTTGGCGTTGAGGAGCTGCCCAATCGGTTACA 660
Db 601 TTGACAGATATTCGCTATGATAATGCTTTTGGCGTTGAGGAGCTGCCCAATCGGTTACA 660
```

QY 661 GGTAAATATCTTTGTTGCTCAGGTCGGCTTAGCTTTACAGACGAGGTGGTGTTCCT 720  
 DB 661 GGTAAATATCTTTGTTGCTCAGGTCGGCTTAGCTTTACAGACGAGGTGGTGTTCCT 720  
 QY 721 AACATAGATAGATACATCAACAGACGCTTCCCTGGGTATTCCTGTAAGTATGGTGATGAC 780  
 DB 721 AACATAGATAGATACATCAACAGACGCTTCCCTGGGTATTCCTGTAAGTATGGTGATGAC 780  
 QY 781 AGGTGCTTGACCAACTATGCAACTGATTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 840  
 DB 781 AGGTGCTTGACCAACTATGCAACTGATTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 840  
 QY 841 TGATATACAGATGTTCCCTGACAAGATGCTTACTTGAAGCAGCAAAACCGCTGGAC 900  
 DB 841 TGATATACAGATGTTCCCTGACAAGATGCTTACTTGAAGCAGCAAAACCGCTGGAC 900  
 QY 901 AGTCCCTTTTAGAGAGTCCATTTATTCGTTTGAAGAAATCATGAACAATCCTTTGTA 960  
 DB 901 AGTCCCTTTTAGAGAGTCCATTTATTCGTTTGAAGAAATCATGAACAATCCTTTGTA 960  
 QY 961 GGCCTATGAGCACTACTGAGGTGCTATGTTTATGATGCTGTTTATTCCTGCTGGAT 1020  
 DB 961 GGCCTATGAGCACTACTGAGGTGCTATGTTTATGATGCTGTTTATTCCTGCTGGAT 1020  
 QY 1021 TTCTTTGAGCAATGTCAGAAATTTGATTGGCTCAGGGTTTATGCTTCTGCTGATT 1080  
 DB 1021 TTCTTTGAGCAATGTCAGAAATTTGATTGGCTCAGGGTTTATGCTTCTGCTGATT 1080  
 QY 1081 ATCTTCATGTTGCTGCTGTCGGAACATTCATTATGATGCTTGAAGCAGCCTGCTCCTC 1140  
 DB 1081 ATCTTCATGTTGCTGCTGTCGGAACATTCATTATGATGCTTGAAGCAGCCTGCTCCTC 1140  
 QY 1141 TTGTTATCTCGTTTATGGGTGCTGCTATGTTTGTCTACAGCCTTGAAGTATAT 1200  
 DB 1141 TTGTTATCTCGTTTATGGGTGCTGCTATGTTTGTCTACAGCCTTGAAGTATAT 1200  
 QY 1201 TCTCTTTTACTATAGAAATGCTGAGTGGGACACGTAAGAAATATTATATA 1254  
 DB 1201 TCTCTTTTACTATAGAAATGCTGAGTGGGACACGTAAGAAATATTATATA 1254

RESULT 3

US-10-011-771B-1  
 ; Sequence 1, Application US/10011771B  
 ; Publication No. US20030082780A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weigel, Paul H.  
 ; APPLICANT: DeAngelis, Paul  
 ; APPLICANT: Kumari, Kshama  
 ; TITLE OF INVENTION: Hyaluronan Synthase Gene and Uses Thereof  
 ; FILE REFERENCE: 3554.011  
 ; CURRENT APPLICATION NUMBER: US/10/011,771B  
 ; CURRENT FILING DATE: 2001-10-11  
 ; PRIOR APPLICATION NUMBER: US 09/178,851  
 ; PRIOR FILING DATE: 1998-10-26  
 ; PRIOR APPLICATION NUMBER: US 60/064,435  
 ; PRIOR FILING DATE: 1997-10-31  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 1254  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus equisimilis  
 US-10-011-771B-1

Query Match 100.0% Score 1254; DB 14; Length 1254;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGAGACATTAATAAACCTCAATACTGTTGGCTTTAGTATTTTGGGTACTGTG 60  
 DB 1 ATGAGACATTAATAAACCTCAATACTGTTGGCTTTAGTATTTTGGGTACTGTG 60

QY 61 ATTACGTCAATGTTATCTCTTTGGTCTAAAGAGCTTGTCAATTTATGCTTTTG 120  
 DB 61 ATTACGTCAATGTTATCTCTTTGGTCTAAAGAGCTTGTCAATTTATGCTTTTG 120  
 QY 121 CTGATAGCTTACCTATTAGTCAAAATGTCCTATCTCTTTTACAAGCCATTAAAGGA 180  
 DB 121 CTGATAGCTTACCTATTAGTCAAAATGTCCTATCTCTTTTACAAGCCATTAAAGGA 180  
 QY 181 AGGCTGGGCAATATAAGGTGCGCCATTAATCCCTCTTATTAAGAGATGCTGATCA 240  
 DB 181 AGGCTGGGCAATATAAGGTGCGCCATTAATCCCTCTTATTAAGAGATGCTGATCA 240  
 QY 241 TTGCTAGAGACCTTAAAGAGTTCAGCAGCAAACTATCCCTTAGCAGAAATTTATGT 300  
 DB 241 TTGCTAGAGACCTTAAAGAGTTCAGCAGCAAACTATCCCTTAGCAGAAATTTATGT 300  
 QY 301 GTTGACGATGGAAGTGTGATGAGACAGGTATTAAAGCATTGAAGACTATGCGTGAC 360  
 DB 301 GTTGACGATGGAAGTGTGATGAGACAGGTATTAAAGCATTGAAGACTATGCGTGAC 360  
 QY 361 ACTGGTGACCTATCAAGCAATGTCATTGTCATCGTCAGAGAAAATCAAGAAAGCGT 420  
 DB 361 ACTGGTGACCTATCAAGCAATGTCATTGTCATCGTCAGAGAAAATCAAGAAAGCGT 420  
 QY 421 CATTGACAGGCTGGGCTTTGAAGATCAGACGCTGATGCTCTTTTGCAGGTTGACTCA 480  
 DB 421 CATTGACAGGCTGGGCTTTGAAGATCAGACGCTGATGCTCTTTTGCAGGTTGACTCA 480  
 QY 481 GATACTTATATCTACCTGATGCTTTTAGAGAGTGTAAAGACCTTTAATGACCAACT 540  
 DB 481 GATACTTATATCTACCTGATGCTTTTAGAGAGTGTAAAGACCTTTAATGACCAACT 540  
 QY 541 GTTTTGTGCGACGGGTCACCTTAATGTCAGAAATAGACAAACCAATCTTTAACACCG 600  
 DB 541 GTTTTGTGCGACGGGTCACCTTAATGTCAGAAATAGACAAACCAATCTTTAACACCG 600  
 QY 601 TTGACAGATATGCTGATGATAATGCTTTTGGCGTTGAACGAGCTGCCCAATCGCTTACA 660  
 DB 601 TTGACAGATATGCTGATGATAATGCTTTTGGCGTTGAACGAGCTGCCCAATCGCTTACA 660  
 QY 661 GGTAAATATCTTTGCTCAGGTCGGCTTAGCTTTACAGACGAGGTGGTGTTCCT 720  
 DB 661 GGTAAATATCTTTGCTCAGGTCGGCTTAGCTTTACAGACGAGGTGGTGTTCCT 720  
 QY 721 AACATAGATAGATACATCAACAGACCTTCCCTGGGTATTCCTGTAAGTATGGTGATGAC 780  
 DB 721 AACATAGATAGATACATCAACAGACCTTCCCTGGGTATTCCTGTAAGTATGGTGATGAC 780  
 QY 781 AGGTGCTTGACCAACTATGCAACTGATTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 840  
 DB 781 AGGTGCTTGACCAACTATGCAACTGATTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 840  
 QY 841 TGATATACAGATGTTCCCTGACAAGATGCTTACTTGAAGCAGCAAAACCGCTGGAC 900  
 DB 841 TGATATACAGATGTTCCCTGACAAGATGCTTACTTGAAGCAGCAAAACCGCTGGAC 900  
 QY 901 AAGTCCCTTTTAGAGAGTCCATTTATTCGTTTGAAGAAATCATGAACAATCCTTTGTA 960  
 DB 901 AAGTCCCTTTTAGAGAGTCCATTTATTCGTTTGAAGAAATCATGAACAATCCTTTGTA 960  
 QY 961 GGCCTATGAGCACTACTGAGGTGCTATGTTTATGATGCTGTTTATTCCTGCTGGAT 1020  
 DB 961 GGCCTATGAGCACTACTGAGGTGCTATGTTTATGATGCTGTTTATTCCTGCTGGAT 1020  
 QY 1021 TTCTTTGAGCAATGTCAGAAATTTGATTGGCTCAGGGTTTATGCTTCTGCTGATT 1080  
 DB 1021 TTCTTTGAGCAATGTCAGAAATTTGATTGGCTCAGGGTTTATGCTTCTGCTGATT 1080  
 QY 1081 ATCTTCATGTTGCTGCTGTCGGAACATTCATTATGATGCTTGAAGCAGCCTGCTCCTC 1140  
 DB 1081 ATCTTCATGTTGCTGCTGTCGGAACATTCATTATGATGCTTGAAGCAGCCTGCTCCTC 1140

QY	481	GATACCTATATCTACCGCTGATGCTTTAGAGAGTGTGTTAAAAACCTTTAATGACCCCACT	540
Db	481	GATACCTATATCTACCGCTGATGCTTTAGAGAGTGTGTTAAAAACCTTTAATGACCCCACT	540
QY	541	GTGTTTGTGCGAGGGTCACCTTTATGTCAGAAATAGACAACCAATCTCTTTACAGCG	600
Db	541	GTGTTTGTGCGAGGGTCACCTTTATGTCAGAAATAGACAACCAATCTCTTTACAGCG	600

601	TTGACAGATATTGCTATGTGTAATGCTTTTGGCTTGACAGAGCTGCCCATCCGTACA	660
601	GGTAATATCCCTGTTTGTCTCAGGTCGCGCTTACGCTTTACAGAGCGAGGTCGTTGTCT	720
661	GGTAATATCCCTGTTTGTCTCAGGTCGCGCTTACGCTTTACAGAGCGAGGTCGTTGTCT	720
721	AACATAGATAGATACATCAACACAGACCTTCCTGGGTATTCCTGAAGTATTGGTGATAC	780
721	AACATAGATAGATACATCAACACAGACCTTCCTGGGTATTCCTGAAGTATTGGTGATAC	780

1021	TCCTCTGAGGCAATGTCAGAGAAATTTGATATGCTCTCAGGGGTTTTAGCCTCTCTCTGCGAATTT	1080
QY		
1021	TTCTTTGTAGGCAATGTCAGAGAAATTTGATATGCTCTCAGGGGTTTTAGCCTCTCTCTGCGAATTT	1080
Db		
1021	TTCTTTGTAGGCAATGTCAGAGAAATTTGATATGCTCTCAGGGGTTTTAGCCTCTCTCTGCGAATTT	1080
QY		
1081	ATCTTCATTTGTGCCCTGTGTCGGGAACATTCATACATGCTTTAAGCACCCGCTGCTCTCTC	1140
QY		
1081	AICTTCATTTGTGCCCTGTGTCGGGAACATTCATACATGCTTTAAGCACCCGCTGCTCTCTC	1140
Db		
1141	TTGCTATCTCGTTTTATAGGGGTGCTGCATTTGTTTGTCTACAGCCCTTGAATTAATATAT	1200
QY		
1141	TTGCTATCTCGTTTTATAGGGGTGCTGCATTTGTTTGTCTACAGCCCTTGAATTAATATAT	1200
Db		
1201	TCCTCTTTTACTATTAGAAATGCTGACTGGGGGAACACGTAAAAAAATTTATTATTA	1254
QY		
1201	TCCTCTTTTACTATTAGAAATGCTGACTGGGGGAACACGTAAAAAAATTTATTATTA	1254
Db		

5,185  
644

; LENGTH: 1251  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus equisimilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1251)  
 ; OTHER INFORMATION:  
 US-10-326-185-1

Query Match 97.7%; Score 1225.4; DB 12; Length 1251;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1235; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 ATGAGACATTAATAAAGCTTATCTGTTGGCCCTTATGATTTTGGTACCTGTTG 60  
 Db 1 ATGAGACATTAATAAAGCTTATCTGTTGGCCCTTATGATTTTGGTACCTGTTG 60  
 QY 61 ATTACGCTCAATGTTATCTCTTTGGTCTAAAGAGCTTGTCAATTTATGCTTTTG 120  
 Db 61 ATTACGCTCAATGTTATCTCTTTGGTCTAAAGAGCTTGTCAATTTATGCTTTTG 120  
 QY 121 CTGATAGCTTACCTTATGCTCAATGCTTATCTCTTTTAAAGCCATTTAAGGGA 180  
 Db 121 CTGATAGCTTACCTTATGCTCAATGCTTATCTCTTTTAAAGCCATTTAAGGGA 180  
 QY 181 AGGGCTGGCAATATAAGGTTGAGCCATTAATCCCTCTTATAAGAGATGCTGATCA 240  
 Db 181 AGGGCTGGCAATATAAGGTTGAGCCATTAATCCCTCTTATAAGAGATGCTGATCA 240  
 QY 241 TTGCTAGAGACCTTAAAGTGTTCAGCAGCAACCTATCCCTTAGCAGAAATTTATGTT 300  
 Db 241 TTGCTAGAGACCTTAAAGTGTTCAGCAGCAACCTATCCCTTAGCAGAAATTTATGTT 300  
 QY 301 GTTGACGATGAAGTGTCTGATGAGCAGAGTATTAAGCCGATTGAGGCTGCTGAC 360  
 Db 301 GTTGACGATGAAGTGTCTGATGAGCAGAGTATTAAGCCGATTGAGGCTGCTGAC 360  
 QY 361 ACTGCTGACCTTCAAGCAATGCTATGTTTCAGGAGTGTAAAGCAATCAAGGAGGCT 420  
 Db 361 ACTGCTGACCTTCAAGCAATGCTATGTTTCAGGAGTGTAAAGCAATCAAGGAGGCT 420  
 QY 421 CATGACAGAGCTTGGGCTTTCAAGATGATGAGCTGATGCTTTTGGACCTGATCTCA 480  
 Db 421 CATGACAGAGCTTGGGCTTTCAAGATGATGAGCTGATGCTTTTGGACCTGATCTCA 480  
 QY 481 GATATATATCTACCTGATGCTTTAGAGGAGTGTAAAGCAATCAAGGAGGCT 540  
 Db 481 GATATATATCTACCTGATGCTTTAGAGGAGTGTAAAGCAATCAAGGAGGCT 540  
 QY 541 GTTTTCTGCGAGGGTCACTTAATGTCAGAAATAGACAAACCAATCTTTAACAGC 600  
 Db 541 GTTTTCTGCGAGGGTCACTTAATGTCAGAAATAGACAAACCAATCTTTAACAGC 600  
 QY 601 TTGACAGATATTCGCTATGATAATGCTTTTGGCTTGAAGCAGCTGCCAATCCGTACA 660  
 Db 601 TTGACAGATATTCGCTATGATAATGCTTTTGGCTTGAAGCAGCTGCCAATCCGTACA 660  
 QY 661 GGTATATTCGCTTGTGCTCAGGCGGCTTACAGCGAGGCTGCTGCTTCCCT 720  
 Db 661 GGTATATTCGCTTGTGCTCAGGCGGCTTACAGCGAGGCTGCTGCTTCCCT 720  
 QY 721 AACATAGATAGATACATCAACGACCTTCCCTGGGATTCCTGTAAGTATGTTGATGAC 780  
 Db 721 AACATAGATAGATACATCAACGACCTTCCCTGGGATTCCTGTAAGTATGTTGATGAC 780  
 QY 781 AGGTGCTTGACCAATATGCACTGATTTAGGAAAGAGCTTTTATCAATCCACTGCTAAA 840  
 Db 781 AGGTGCTTGACCAATATGCACTGATTTAGGAAAGAGCTTTTATCAATCCACTGCTAAA 840  
 QY 841 TGTATTTACAGATGTTCTGACAGATGCTACTTTACTTTGAAGCAGCAAAACCGTGGAC 900  
 Db 841 TGTATTTACAGATGTTCTGACAGATGCTACTTTACTTTGAAGCAGCAAAACCGTGGAC 900

RESULT 6

US-10-326-185-108  
 ; Sequence 108, Application US/10326185  
 ; Publication No. US20030175902A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sloma, Alan  
 ; APPLICANT: Behr, Regine  
 ; APPLICANT: Widner, William  
 ; APPLICANT: Tang, Maria  
 ; APPLICANT: Sternberg, David  
 ; APPLICANT: Brown, Stephen  
 ; TITLE OF INVENTION: Methods for Producing Hyaluronan In a Recombinant Host Cell  
 ; FILE REFERENCE: 10241.200-US  
 ; CURRENT APPLICATION NUMBER: US/10/326,185  
 ; PRIOR FILING DATE: 2002-12-20  
 ; PRIOR APPLICATION NUMBER: US 60/342,644  
 ; NUMBER OF SEQ ID NOS: 108  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 108  
 ; LENGTH: 5158  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus equisimilis  
 US-10-326-185-108

Query Match 91.98; Score 1152; DB 12; Length 5158;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1152; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 TCAATTAATGCTTTTGTGCTGATGCTTACCTATAGTCAAAATGCTCTTATCCCTTTT 162  
 Db 1 TCAATTAATGCTTTTGTGCTGATGCTTACCTATAGTCAAAATGCTCTTATCCCTTTT 60  
 QY 163 TACAGCCATTTAAGGAGGCTGGGCAATATAAGTTGACGCAATATTCCTCTTAT 222  
 Db 61 TACAGCCATTTAAGGAGGCTGGGCAATATAAGTTGACGCAATATTCCTCTTAT 120  
 QY 223 AACGAGATGCTGAGTCATGCTAGACCTTAAAGTGTTCAGCAGCAAAACCTATCCC 282  
 Db 121 AACGAGATGCTGAGTCATGCTAGACCTTAAAGTGTTCAGCAGCAAAACCTATCCC 180  
 QY 283 CTAGCAGAAATTTATGTTGCTGATGGAAGTGTCTGATGAGCAGGATTAAGCGCAT 342  
 Db 181 CTAGCAGAAATTTATGTTGCTGATGGAAGTGTCTGATGAGCAGGATTAAGCGCAT 240  
 QY 343 GAAGACTATGCGTGACACTGGTACCTATCAAGCAATGCTATGTCATCGGTCAGAG 402  
 Db 241 GAAGACTATGCGTGACACTGGTACCTATCAAGCAATGCTATGTCATCGGTCAGAG 300

QY	403	AAAAATCAAGGAAAGCGTCATGCACAGGSCCTGGGCGCTTTGAAAGATCAGACGCGTGAATGC	466
DB	301	AAAAATCAAGGAAAGCGTCATGCACAGGSCCTGGGCGCTTTGAAAGATCAGACGCGTGAATGC	360
QY	463	TTTTTGCACCGTGTGACTCAGATACTATATCTACCCGTGATGCTTTAGAGAGTGTGTTAAAA	522
DB	361	TTTTTGCACCGTGTGACTCAGATACTATATCTACCCGTGATGCTTTAGAGAGTGTGTTAAAA	420
QY	523	ACCTTTAATGACCCAACTGTTTTGCTGCGAGGGTGACCTTAATGTCAGAAAAATAGACAA	582
DB	421	ACCTTTAATGACCCAACTGTTTTGCTGCGAGGGTGACCTTAATGTCAGAAAAATAGACAA	480
QY	583	ACCAATCTCTTAAACAGCGTTGACAGATATTCGCGTATGATAATGCTTTGGCGTTGAACGA	642
DB	481	ACCAATCTCTTAAACAGCGTTGACAGATATTCGCGTATGATAATGCTTTGGCGTTGAACGA	540
QY	643	GCTGCCAATCCGTTACAGGTAATATCCTTTGTTGCTCAGGTCGCGTTAGCGTTTACAGA	702
DB	541	GCTGCCAATCCGTTACAGGTAATATCCTTTGTTGCTCAGGTCGCGTTAGCGTTTACAGA	600
QY	703	CGGAGGTGGTGTTCCTTAACATGATAGATACATCAACCAGACCTTCCTGGGTATTCTCT	762
DB	601	CGGAGGTGGTGTTCCTTAACATGATAGATACATCAACCAGACCTTCCTGGGTATTCTCT	660
QY	763	GTAAGTATTGGTGATGACAGGTGCTTCACCACTATGCACTGATTTAGGAAGACTGTT	822
DB	661	GTAAGTATTGGTGATGACAGGTGCTTCACCACTATGCACTGATTTAGGAAGACTGTT	720
QY	823	TATCAATCCACTGCTAAATGTATTACAGATGTCTCGACAAGATGTCATCTTACTTGAAG	882
DB	721	TATCAATCCACTGCTAAATGTATTACAGATGTCTCGACAAGATGTCATCTTACTTGAAG	780
QY	883	CAGCAAAACCGCTGGACAGTCTCTTTAGAGAGTCCATATTCTTGTTTAAAGAAATC	942
DB	781	CAGCAAAACCGCTGGACAGTCTCTTTAGAGAGTCCATATTCTTGTTTAAAGAAATC	840
QY	943	ATGAACAATCCCTTTGTAGCCCTATGGACCACTATGAGGTGTCATGTTTATGATGCTT	1002
DB	841	ATGAACAATCCCTTTGTAGCCCTATGGACCACTATGAGGTGTCATGTTTATGATGCTT	900
QY	1003	GTTTATCTGTGTGGATTTCTTTGTAGCCAAATGTCAGAGATTTGATGGCTCAGGGTT	1062
DB	901	GTTTATCTGTGTGGATTTCTTTGTAGCCAAATGTCAGAGATTTGATGGCTCAGGGTT	960
QY	1063	TTAGCCCTTTGCGTGAATATCTTCATGTTGCGCTGTGTCGGAACATTCATTCATGCTT	1122
DB	961	TTAGCCCTTTGCGTGAATATCTTCATGTTGCGCTGTGTCGGAACATTCATTCATGCTT	1020
QY	1123	AGGACCGGTGCCCTTCCTGTTATCTCCGTTTATGGGTGCTGCATTTGTTTGCCTTA	1182
DB	1021	AGGACCGGTGCCCTTCCTGTTATCTCCGTTTATGGGTGCTGCATTTGTTTGCCTTA	1080
QY	1183	CAGCCCTTGAATATATTCTCTTTTTACTATTAAGAAATCTGACTGGGGAACACAGTAAA	1242
DB	1081	CAGCCCTTGAATATATTCTCTTTTTACTATTAAGAAATCTGACTGGGGAACACAGTAAA	1140
QY	1243	AAATATATTATAA	1254
DB	1141	AAATATATTATAA	1152

## RESULT 7

US-102-326-185-102  
/ Sequence 102, Application US/10326185  
/ Publication No. US20030175902A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Sloma, Alan  
/ APPLICANT: Behr, Regine  
/ APPLICANT: Widner, William  
/ APPLICANT: Tang, Maria  
/ APPLICANT: Sternberg, David  
/ APPLICANT: Brown, Stephen



Qy	1201	TCTCTTTTACTATTAGAAATGCTGACTGGGGAACACGTA	1244
Db	1220	TGGTATTTCATTAGAAATGCTACATGGGGAACTCGTA	1263

## RESULT 9

```

US-10-172-527-13
:
: Sequence 13, Application US/10172527
: Publication No. US20030092118A1
: GENERAL INFORMATION:
: APPLICANT: Weigel, Paul H
: APPLICANT: Kumari, Kshama
: APPLICANT: Derangelis, Paul
: TITLE OF INVENTION: HYALURONAN SYNTHASE G
: FILE REFERENCE: 3554.048
: CURRENT APPLICATION NUMBER: US/10/172,527
: CURRENT FILING DATE: 2002-06-13
: PRIOR APPLICATION NUMBER: 60/297,788
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: 60/297,744
: PRIOR FILING DATE: 2001-06-13
: PRIOR APPLICATION NUMBER: 09/469,200
: PRIOR FILING DATE: 1999-12-21
: PRIOR APPLICATION NUMBER: 09/178,851
: PRIOR FILING DATE: 1998-10-26
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 13
: LENGTH: 1440
: TYPE: DNA
: ORGANISM: Streptococcus pyogenes
US-10-172-527-13

```

Query Match	48.2%;	Score	604.6;	DB	14;	Length	1440;
Best Local Similarity	68.7%;	pred. No.	8.8e-162;				
Matches	848;	Conservative	0;	Mismatches	384;	Indels	3;
							Gaps
							1;
QY	10	TTAAAAACCTCATACCTTGTGGCTTTAGTATTTTTGGGTACTCTTGATTACGTC	69				
Db	130	TTTAAAAAAACTTTAATGTGTTATCCCTTATTTTGTATATCATCTTGATTATCTA	189				
QY	70	AATCTTATCTCTTTGGTGCTAAGAACCTGTCAATTTATGGCTTTTGGCTGATAGCT	129				
Db	190	AATATGTAICTTTTGGAAAC---ATCACTAGGAATTTATGGAGTAATATTAATAACCC	246				
QY	130	TACCTATTAGTCAAAATGCTCTTATCCCTTTTTTTTACAACCATTTTATAGCAACAGCGGACCC	180				

[illegible]

Qy	550	GCAGCGGGTCACCTTAATGTGCAAGAAATAGACAAACCAATCTCTTAAACAGCGCTTGACAGAT	609
Db	667	GCAACAGGACATTTTGAATGCTAGAAACAGACAAACTAATCTATTAAACGGACCTTACAGAT	726
Qy	610	ATTTCGCTATGATTAATGCTTTGGCGTTTGAACGAGCTGCCAATCCGTTACAGGTAATATC	669
Db	727	ATCCGTTTACGTAATGCTTTGGGTGGAGCGTGTCTCAATCATTTACAGGTAATATT	786
Qy	670	CTTGTGTCAGTCCGCTTACGCTTTTACAGACGCGAGGTGTTCTCTCAACATAGAT	729
Db	787	TTAGTTTGTCTCAGGACCATTGAGTATTTATTCGAGCTGAAGTGATTAATCTCACTACTAGAG	846
Qy	730	AGATACATCAACGACGCTTCCTCGGTATTCCTGTGTAAGTATTTGGTGATGACAGGTGCTTG	789
Db	847	CGTATAAAAAATCAACATCTCTAGTTTACCTGTTTAGCATTTGGGATGCGATGATGTTAA	906
Qy	790	ACCACTATGCAATGATTTAGSAAAGACTGTTTATCAATCCACTGCTAAATGATTATACA	849
Db	907	ACAAATTATGCTATTGATTTTAGCGCAGCTGCTACCAATCCACAGCTAGATGTGATCT	966
Qy	850	GAATGTCCTGACAGATGCTTACTTTCTTGAGCAGCAAAACCGCTGGACAAAGTCCTCTC	909
Db	967	GATGTACCTTTCCAAATTAAGAAAGTATTTAAAGACACAAATCGATGGATAAATCTTTT	1026
Qy	910	TTTAGAGAGTCCATTAATTTCTGTTAAGSAAATCATGAACAATCCTTTGTGTAGCCCTATGG	969
Db	1027	TTTAGAGAATCTATTAATTCGTGTTAAAAAATCTTCTTAATCCCATGTTGCCCTTATGG	1086
Qy	970	ACCATACTTGAGTGTCTATGTTTATGAGCTGTGTTTATTCCTGTGGTATTTCTTTGTA	1039
Db	1087	ACTATTTTCGAAGTCGTTATGTTTATCATGTGATTGTCGAATTTGGGAATCTTTTGTTT	1146
Qy	1030	GGCAATGTCAGAAATTTGATTGGCTCAGGCTTTTACCCCTTCCTGGTATTCCTTCATT	1089
Db	1147	AATCAGCTATTCAATTAGACCTTATTAACACTTTTGGCTTTTATCCCATCATCTTTATC	1206
Qy	1090	GTTGCCCTGTGCGGAACATTCATTACATGCTTAAGCACCCGCTGTCTCTTCTTGTTATCT	1149
Db	1207	GTTGCTTTATGTCGTAAATGTTTCATTATATGTTGTCAAACATCCTGCTAGTTTTCCTATCT	1266
Qy	1150	CGGTTTTATGGGTGCTGATCTTCTTTGTCCTACAGCGCTTGAATATATATCTCTCTTTT	1209
Db	1267	CCTCTGTATGGAAATATACACTGTTTGTCCTTACAGCGCCCTAAAACCTTATCTTTATGC	1326
Qy	1210	ACTATTAGAAATCGTGACTGGGGAACACGTAAAA	1244
Db	1327	ACCATTAATAATACGGAATGGGAAATGGGACACCGCTAAAA	1361

RESULT 10

```

1  US-10-124-222-1
2  ; Sequence 1, Application US/10124222
3  ; Publication No. US20030104415A1
4  ; GENERAL INFORMATION:
5  ; APPLICANT: WEIGEL, PAUL E.
6  ; APPLICANT: DEANGELIS, PAUL L.
7  ; APPLICANT: PAPACONSTANTINO, JOHN
8  ; TITLE OF INVENTION: HYALURONATE SYNTHASE GENES AND US55 THEREOF
9  ; FILE REFERENCE: 35541.089
10 ; CURRENT APPLICATION NUMBER: US/10/124,222
11 ; PRIOR FILING DATE: 2002-04-15
12 ; PRIOR APPLICATION NUMBER: 09/146,893
13 ; PRIOR FILING DATE: 1998-09-03
14 ; NUMBER OF SEQ ID NOS: 9
15 ; SOFTWARE: PatentIn version 3.1
16 ; SEQ ID NO 1
17 ; LENGTH: 1512
18 ; TYPE: DNA
19 ; ORGANISM: Streptococcus pyogenes
20 ; FEATURE:
21 ; NAME/KEY: CDS
22 ; LOCATION: (193)..(1449)
23 ; OTHER INFORMATION:

```





Db 667 TTAGTTGGTCCAGGACCATTTAGTATTTATCGAGCTGAAGTATTTCCCTAACCTTAGAG 726  
QY 730 AGATACATCAACAGACCTTCCTGGGTATTCCTGTAGTATTTGGTATGACAGAGTGTG 789  
Db 727 CGGTATATAAATCAACATTCCTAGGTTTACCTGTAGCATTTGGGATGATCGATGTTTA 786  
QY 790 ACCAATATGCAACTGATTTAGGAAGACTGTTTATCAATCCACCTGCTAAATGATTTACA 849  
Db 787 ACARATATGCTATTTAGTATTTAGGACGCTGTCTACCATCAACAGCTAGATGATATC 846  
QY 850 GATGTTCCGCAAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAACAGTCTCTTC 909  
Db 847 GATGCTACCTTCCCAATTAAGTATTTAAAGCAACAAAATCGATGGAATTAATCTTT 906  
QY 910 TTAGAGAGTCCATATTTCTGTAGAAATCATGAACAACTCTTTTGTAGCCCTATGG 969  
Db 907 TTPAAGAATCATTTATTTCTGTAAAGAAATCTTTCTAATCCCATCGTTGCTTATGG 966  
QY 970 ACCAATATGAGGTGCTATGTTTATGATGCTGTTTATCTGTGCTGATTTCTTTGTA 1029  
Db 967 ACTATTTGCGAGTCGTATGTTTATGATGTTGATGTTGCGCAATTTGGGAATCTTTGTT 1026  
QY 1030 GGCATATGACAGAAATTTGATGCGCTCAGGCTTTTACGCTTTCTGCTGATTTCTTCATT 1089  
Db 1027 AATCAAGCTATTCAATTAGACCTTATTAACCTTTTGGCTTTTATCCATCATCTTTATC 1086  
QY 1090 GTTGCCCTGCTGCGAATCATTTACATGCTTAAGCAGCCGCTCTCTCTCTTTGTTATCT 1149  
Db 1087 GTTGCTTATGCTGATGTTTATATATGATCAATCAATCATCTGCTAGTTTGTGTTATCT 1146  
QY 1150 CCGTTTATGCGGTGCTGCAATTTGTTGCTCCTCAGCCCTTGAAATTTATATCTCTTTT 1209  
Db 1147 CCTCTATGGAATATTACACTGTTTGTCTTACAGCCCTAAACCTTTATCTTTATGC 1206  
QY 1210 ACTATTAGAAATGCTGACTGCGGGAACACGTTAAAA 1244  
Db 1207 ACCATTAAAAATAGGAATGGGGAACACGTTAAAA 1241

RESULT 13  
US-10-172-527-17  
; Sequence 17, Application US/10172527  
; Publication No. US20030092118A1  
; GENERAL INFORMATION:  
; APPLICANT: Weigel, Paul H  
; APPLICANT: Kumari, Kshama  
; APPLICANT: DeAngelis, Paul  
; TITLE OF INVENTION: HYALURONAN SYNTHASE GENES AND EXPRESSION THEREOF IN BACILLUS SUBT  
; FILE REFERENCE: 3554,048  
; CURRENT APPLICATION NUMBER: US/10/172,527  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: 60/297,788  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: 60/297,744  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: 09/469,200  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 09/178,851  
; PRIOR FILING DATE: 1998-10-26  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 17  
; LENGTH: 1200  
; TYPE: DNA  
; ORGANISM: Bacillus anthracis px01  
US-10-172-527-17  
Query Match 22.6%; Score 282.8; DB 14; Length 1200;  
Best Local Similarity 57.8%; Pred. No. 5.9e-70;  
Matches 503; Conservative 0; Mismatches 367; Indels 0; Gaps 0;  
QY 383 TCATGTTTCATCGTCCAGAGAAAATCAAGAAAGCTCATCCACAGCGCTGGCGCTTGG 442

Db 255 TTAATCGTACATCGCTTTACCTAAGAAATTCGCGGAAAGACATGCTCAATATATGGCTTTA 314  
QY 443 AAAGATCAGAGCGCTGATGTCCTTTTGTACCGCTTGACTCAGATCTTATATCTACCTGATG 502  
Db 315 AACGGAACAACAGCAGATGCTATTGTACCATTTGATTTGATGTTGATTTGTTCCCAATG 374  
QY 503 CTTTAGAGGAGTTGTTTAAAAACCTTTTAAAGACCAACTGTTTGTGCGGAGCGGTCAAC 562  
Db 375 CTGTATAGAGAGTTATTGAAACCCCTTTAATGATGAAAAAGTAAATGGCCCAACATGCTCAG 434  
QY 563 TTAATGTCAGAAATAGACAAACCAATCTCTTACACGCTTGCAGATATTTCCCTATGATA 622  
Db 435 TGAACATTCGTAATAGAAATGATTAATTTAACAAAACTAATTTGATATGCGTTATGACA 494  
QY 623 ATGCTTTTGGCGTTGACAGAGCTGCCCAATTCGCTTACAGGTAATATCTCTTTGCTCAG 682  
Db 495 ATGCGTTCCGTTGGAGCGTGCAGCAGCTCCGTAACAGGAATGTTCTTGTGTTAGTG 554  
QY 683 GTCCGCTTACGTTTACAGACGCGAGGTGTTCTTCCCTAACATAGATAGATACATCAACC 742  
Db 555 GCGCGTTAAGTTGTTATCTGTAGAGAGTAATACTGAAAAATTTAGAACATTTATGAAAGTC 614  
QY 743 AGACCTTCTCGGTATTTCTCTGTAAGTATTTGGTATGACAGGCTGTTGACCAACTATGCAA 802  
Db 615 AGATGTTCTTGGTGAGGAGGTGCGAGTTGGAGATGATAGATGCTAACATAATATGCTA 674  
QY 803 CTGATTTAGAAAGACTGTTTATCAATCCACTGCTAAATGTTATTTACAGATGTTCTCGACA 862  
Db 675 TTTTGAAGGAAACACAGTTTATCAATCCACTGCTCGATGTTACTGATGCTCCAATA 734  
QY 863 AGATGCTCTACTTACTTGAAGCAGCAAAACCGCTGGAACAAGTCTCTTTAGAGAGTCCA 922  
Db 735 CATTAACAACTTTCTTAAACAGCAACTAGCTTGGAAACAAGTCAATTTTGTAGAGAAAGTT 794  
QY 923 TTAATTTCTGTTAAGAAAAATCATGAACATCTTTTGTAGCCCTATGACCACTACTTGAGG 982  
Db 795 TAATTTCACTTGGCATTTGTTATGAAAAACCAATGTTCTTGTGTTGACAAATTTTCCAAA 854  
QY 983 TGTCTATGTTTATGATGCTTTTATTTCTGTGTTGTTGATTTCTTTGTAGGCAATGTCAGAG 1042  
Db 855 TATCGTTATGGAATTTTATTTGGGCTTTCCCTACTTCTAAGTATTAATTTCTCAAGGCAAGTC 914  
QY 1043 AATTTGATTCGCTCAGGCTTTTACCTTTCTGTCATTTCTTCATTTGTTGCCCTGTGTC 1102  
Db 915 ATGTAGGTTAATTTTGGCTGTTTATTTATTTGGGTTATTTATTTAGCTGTATATGCTA 974  
QY 1103 GGAACATTTCAATCATGCTTAAAGCACCCGCTGCTCTTCTTGTATCTCGTTTATGCGG 1162  
Db 975 GAAATGATTTTATCTATTATAAACATCCCTTACTTCTTACGCGCCATATATGAA 1034  
QY 1163 TCGTGCATTTTGTCTCCTCAGCCCTTGAATTTATTTCTCTTTTACTATTAATAAGT 1222  
Db 1035 TTTCCCATGTTAGCACTATTACCTATACGCTTTTATGCTTTTACTAACTAATAATCTA 1094  
QY 1223 CTGACTGGGGAACACCACTAAAAAATTAAT 1252  
Db 1095 ATGCTGGGGAACACGTTAATTACGTAAT 1124

RESULT 14  
US-09-902-939-2  
; Sequence 2, Application US/09902939  
; Publication No. US20030087850A1  
; GENERAL INFORMATION:  
; APPLICANT: Philip Dehazya  
; APPLICANT: Weillam Chen  
; TITLE OF INVENTION: GENE THERAPY FOR DRY EYE SYNDROME  
; FILE REFERENCE: 2055/0H020-US0  
; CURRENT APPLICATION NUMBER: US/09/902,939  
; CURRENT FILING DATE: 2001-07-10  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2  
; LENGTH: 1659  
; TYPE: DNA  
; ORGANISM: mouse  
US-09-902-939-2

Query Match  
Best Local Similarity 49.1%; Score 61.6; DB 11; Length 1659;  
Matches 222; Conservative 0; Mismatches 224; Indels 6; Gaps 2;  
QY 474 TGACTCAGATATATATACCTGATGCTTTAGAGAGTGTAAACCTT---TAA 530  
DB 633 TGACTCAGATATATATACCTGATGCTTTAGAGAGTGTAAACCTT---TAA 530  
QY 531 TGACCAACTGTTTTTGGCTGAGGCTGCTTAAATGTCAGAAATAGACAACTTCT 590  
DB 693 AGACCCATGCTTGGAGGCTGTTGGAGGAGATGTCAGATTTTAAACAGTATGATTCCTG 752  
QY 591 CTTACAGGCTTGACAGATATTCGCTATGATGCTTTTGGGCTTTGACGAGCTGCCA 650  
DB 753 GATCTCTTCTTCAGCAGCGTGAGATACCTGATGCTTTTAAATATAGAAAGGCTGCCA 812  
QY 651 ATCCCTACAGGTAATATCTTGTTCCTCAGCTCGCTTACGCTTTACAGACGCGAGT 710  
DB 813 GTCATATTTTGGCTGCTCAGTGATGCTTACAGGCTGCTTGGAAATGTACAGAACTCCTT 872  
QY 711 GCTTGTTCCTTAAATATGATATACATCAACAGACCTTCTCGGCTATTCCTGTAAGTAT 770  
DB 873 GCTGATGAATTTTGGAGAGCTGGTACATCAGGAATTCATGGTAACTGATGCTTTT 932  
QY 771 TGGTCATGACAGTGTGACCAACTATGCACTGATTTAGGAAGAGCTGTTTATCAATC 830  
DB 933 TGGTGACGACGACCACTTACCAAGGCTGTTGAGTCTGGGCTATGCAACTAAATACAC 992  
QY 831 CACTG---CTAAATGTTATACAGATGTTCTGACAGATGCTTACTTACTTGAAGCAGA 887  
DB 993 GCTCGGTCCAAGTGCCTTACTGAACTCCCATAGATATCTGAGATGGCTGAACGCA 1052  
QY 888 AAACCGCTGGAACAAGTCTCTTTTAGAGAGT 919  
DB 1053 GACCCGATGGAGCAAGTCTTACTTCCGAGAGT 1084

RESULT 15  
US-10-262-526-3  
; Sequence 3, Application US/10262526  
; Publication No. US20030108531A1  
; GENERAL INFORMATION:  
; APPLICANT: Hideshige Moriya  
; APPLICANT: Yuichi Wada  
; TITLE OF INVENTION: GENES FOR TRANSFECTION INTO BONY TISSUES  
; FILE REFERENCE: OPI418  
; CURRENT APPLICATION NUMBER: US/10/262,526  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: JP 2001-367091  
; PRIOR FILING DATE: 2001-11-30  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 4194  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (508)..(2166)  
US-10-262-526-3

Query Match  
Best Local Similarity 49.1%; Score 61.6; DB 14; Length 4194;  
Matches 222; Conservative 0; Mismatches 224; Indels 6; Gaps 2;  
QY 474 TGACTCAGATATATATACCTGATGCTTTAGAGAGTGTAAACCTT---TAA 530

DB 1140 TGACTCAGATATATATACCTGATGCTTTAGAGAGTGTAAAGTCTTTAGAGCA 1199  
QY 531 TGACCAACTGTTTTTGGCTGAGGCTGCTTAAATGTCAGAAATAGACAACTTCT 590  
DB 1200 AGACCCATGCTTGGAGGCTGTTGGAGGAGATGTCAGATTTTAAACAGTATGATTCCTG 1259  
QY 591 CTTAACACGCTTGACAGATATTCGCTATGATGCTTTTGGGCTTTGACGAGCTGCCA 650  
DB 1260 GATCTCTTCTTCAGCAGCGTGAGATACCTGATGCTTTTAAATATAGAAAGGCTGCCA 1319  
QY 651 ATCCGTTACAGGTAATATCTTGTTCCTCAGCTCGCTTACGCTTTACAGACGCGAGT 710  
DB 1370 GTCATATTTTGGCTGCTCAGTGATGAGGCTTCTGGGAATGTACAGAACTCCTT 1379  
QY 711 GGTGTTCTTAACTATAGATATACATCAACAGACCTTCTCGGCTATTCCTGTAAGTAT 770  
DB 1380 GCTGATGAATTTTGGAGAGCTGGTACATCAGGAATTCATGGTAACTGAGTTT 1439  
QY 771 TGGTGATGACAGGCTGTTGACCACTATGCACTGATTTAGGAAGAGCTGTTTATCAATC 830  
DB 1440 TGGTGACGACGAGGCTTACCAAGGCTGTTGAGTCTGGGCTATGCAACTAAATACAC 1499  
QY 831 CACTG---CTAAATGTTATACAGATGTTCTGACAGATGCTTACTTACTTGAAGCAGA 887  
DB 1500 GCTCGGTCCAAGTGCCTTACTGAACTCCCATAGATATCTGAGATGGCTGAACGCA 1559  
QY 888 AAACCGCTGGAACAAGTCTCTTTTAGAGAGT 919  
DB 1560 GACCCGATGGAGCAAGTCTTACTTCCGAGAGT 1591

Search completed: October 2, 2003, 12:44:53  
Job time : 349 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2003, 09:01:59 ; Search time 2968 seconds  
(without alignments)  
10268.805 Million cell updates/sec

Title: US-09-469-200D-1  
Perfect score: 1254  
Sequence: 1 atgagaacattataaaacacct.....cacgtataaaattattataa 1254

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_esti:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	61.6	4.9	2891	11 AK079729	AK079729 Mus muscu
2	50.6	4.0	1101	29 CNS017WW	AL108602 Drosophil
3	49.4	3.9	809	14 CA324064	CA324064 UI-M-FY0-
4	49.4	3.9	4151	11 AK028582	AK028582 Mus muscu

5	48.4	3.9	1180	13 BX436369	BX436369 BX436369
6	48	3.8	712	13 BX416727	BX416727 BX416727
7	46.8	3.7	712	28 BH980141	BH980141 Odg45g12.
8	46.8	3.7	919	29 CNS01MCA	AL150747 Anopheles
9	46.4	3.7	1702	10 BG165669	BG165669 602345171
10	46.2	3.7	1124	28 B12074	B12074 R25H9-77-2
11	45.8	3.7	976	29 CNS001Q2	AL075095 Drosophil
12	45	3.6	714	13 BX415371	BX415371 BX415371
13	45	3.6	939	29 CNS00CNG	AL059400 Drosophil
14	44.6	3.6	1176	13 BX333445	BX333445 BX333445
15	44.4	3.5	1101	29 CNS00LOO	AL068607 Drosophil
16	44.2	3.5	816	29 CNS02KTV	AL201604 Tetradon
17	44.2	3.5	1201	9 AL531744	AL531744 AL531744
18	44	3.5	485	13 BQ828135	BQ828135 L66in142
19	44	3.5	1101	29 CNS007GT	AL078714 Drosophil
20	43.8	3.5	524	29 CNS00LFT	AL067552 Drosophil
21	43.8	3.5	960	13 BX397886	BX397886 BX397886
22	43.4	3.5	903	29 CNS0011C	AL074609 Drosophil
23	43.4	3.5	1072	29 CNS007OB	AL067418 Drosophil
24	43.2	3.4	643	28 AZ282684	AZ282684 RRCI-23-1
25	43.2	3.4	654	10 BE889709	BE889709 601512740
26	43.2	3.4	945	29 CNS00IKW	AL075212 Drosophil
27	43.2	3.4	1100	13 BX381739	BX381739 BX381739
28	43.2	3.4	1201	9 AL514129	AL514129 AL514129
29	43	3.4	976	13 BX347783	BX347783 BX347783
30	43	3.4	1000	29 CNS00COQ	AL059446 Drosophil
31	42.8	3.4	864	29 CNS016M4	AL106486 Drosophil
32	42.6	3.4	399	29 CNS01609	AL106131 Drosophil
33	42.4	3.4	473	29 BX238763	BX238763 Danio rer
34	42.4	3.4	907	29 CNS021J4	AL176953 Tetradon
35	42.4	3.4	1249	29 BZ566650	BZ566650 pacs2-164
36	42.2	3.4	498	9 AL574975	AL574975 AL574975
37	42.2	3.4	530	13 BX425377	BX425377 BX425377
38	42.2	3.4	883	29 CC261953	CC261953 CH261-184
39	42.2	3.4	913	28 BH155504	BH155504 ENTRS33TR
40	42.2	3.4	915	28 BH139902	BH139902 ENTO549TR
41	42.2	3.4	940	9 AL581799	AL581799 AL581799
42	42.2	3.4	957	28 AZ668597	AZ668597 ENTG038TF
43	42.2	3.4	1472	10 BG421340	BG421340 602451464
44	42	3.3	864	29 CNS0605G	AL407642 T7 end of
45	42	3.3	926	28 BH943155	BH943155 odd64a12.

## ALIGNMENTS

RESULT 1	AK079729	2891 bp	linear	HTC 05-DEC-2002
LOCUS	Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:A430041J01 product:hyaluronan synthase 2, full insert sequence.			
DEFINITION	AK079729			
ACCESSION	AK079729.1	GI:26348186		
VERSION	AK079729.1	GI:26348186		
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)			
MEDLINE	99279253			
PUBMED	10349636			
REFERENCE	1			
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374			
PUBMED	11042159			



```

CNS017WW/c
LOCUS      CNS017WW      1101 bp      DNA      linear      GSS      26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37D07 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL108602
VERSION    AL108602.1  GI:5628906
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1101)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
            - web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelobAC11.

FEATURES             Location/Qualifiers
     source           1..1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN37D07"
                     /clone_lib="DrosBAC"
                     /plasmid="pBelobAC11"
                     /note="end : SP6"

BASE COUNT  533 a 179 c 108 g 125 t 156 others
ORIGIN
Query Match      4.0%; Score 50.6; DB 29; Length 1101;
Best Local Similarity 42.4%; Pred. No. 1.1;
Matches 131; Conservative 22; Mismatches 156; Indels 0; Gaps 0;

QY  907  TTCCTTAGAGTCCATATTCTCTTAGAAGAAATCATGAACAATCCCTTTAGCCCTA 966
      || || || || || || || || || || || || || || || || || || ||
Db  487  TTATKTTGTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 428
      || || || || || || || || || || || || || || || || || || ||

QY  967  TGGACCATCTAGGTCATGTTATGATGCTGTTATCTGCTGGGATTTCTTT 1026
      || || || || || || || || || || || || || || || || || || ||
Db  427  TGTAGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 368
      || || || || || || || || || || || || || || || || || || ||

QY  1027 GTAGCAATGTCAGAAATTTGATTTGGCTCAGGTTTATAGCCCTTCTGGTATTACTTC 1086
      || || || || || || || || || || || || || || || || || || ||
Db  367  TTWTCTTATTCATATTTTGTGTTTATTTTATTTTATTTTATTTTATTTTATTTT 308
      || || || || || || || || || || || || || || || || || || ||

QY  1087 ATGTGTCCTGTCGGAACATTCATTACATGCTTAACACCCGCTGCTCTCTGTTA 1146
      || || || || || || || || || || || || || || || || || || ||
Db  307  TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 248
      || || || || || || || || || || || || || || || || || || ||

QY  1147 TCTCGTTTATGGGTGCTGATTTGCTTCTCTACACCCCTCAAAATATATTCCTTT 1206
      || || || || || || || || || || || || || || || || || || ||
Db  247  TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 188
      || || || || || || || || || || || || || || || || || || ||

QY  1207 TTTACTATT 1215
      || || || || || || || || || || || || || || || || || || ||
Db  187  TTGTGTTATT 179
      || || || || || || || || || || || || || || || || || || ||

RESULT 3
CA324064
LOCUS
DEFINITION UI-M-FY0-cso-g-16-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
            IMAGE: 6822281.5', mRNA sequence.

```

```

ACCESSION  CA324064
VERSION    CA324064.1  GI:24542162
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 809)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cga@phs-research.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pyX-5.
            Location/Qualifiers
     source           1..809
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE: 6822281"
                     /tissue_type="whole brain"
                     /dev_stages="embryo 13.5, 14.5, 16.5, 17.5dpc"
                     /lab_host="DHL0B (T1 phage resistant)"
                     /clone_lib="NIH-BMAP_FY0"
                     /note="Organ: Brain; Vector: pyX- Asc; Site: 1: EcoR I;
                     Site: 2: Not I; The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pyX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     is ACGGAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): Gene Discovery in the
                     Developing Mouse Nervous System", supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                     program coordinator."
BASE COUNT  175 a 218 c 193 g 221 t 2 others
ORIGIN
Query Match      3.9%; Score 49.4; DB 14; Length 809;
Best Local Similarity 49.2%; Pred. No. 1.9;
Matches 159; Conservative 0; Mismatches 161; Indels 3; Gaps 1;

QY  600  CTTGACAGATATTCCTATGATATGCTTTGGCGTTGACGAGCTGCCCAATCCGTTAC 659
      || || || || || || || || || || || || || || || || || || ||
Db  119  CTTGAGCAGTGTGAGTACTGGATGGCTTTCAAGTGGAGCGGGCTCCCACTCTACTT 178
      || || || || || || || || || || || || || || || || || || ||

QY  660  AGGTAATATCTTGTGCTCAGGTCGCCCTTAGCGTTTACACGCGAGGTGGTTGTCC 719
      || || || || || || || || || || || || || || || || || || ||
Db  179  TGGCTGTGTGCAATGATATAGTGGCGCTTTGGGCATGATCCGCAACACCTCTCAGCA 238
      || || || || || || || || || || || || || || || || || || ||

QY  720  TAACATAGATAGATACATCAACCAACCTCTCTGGGTATTTCTTGTAGTATTTGTGATGA 779
      || || || || || || || || || || || || || || || || || || ||
Db  239  GTTCTCGAGGATTGGTACCATCAGAAGTCTCTAGGCAGCAAGTCAGCTTTGGGATCA 298
      || || || || || || || || || || || || || || || || || || ||

QY  780  CAGGTCTCTGACCAACTATGCACTGATTAGGAAA---GACTGTTTATCAATCACTCTGC 836
      || || || || || || || || || || || || || || || || || || ||
Db  299  TCGGCACTTCAACACCGAGTCTGTAGTCTGGCTAGCCGACTTAAGTATACAGACGCTC 358
      || || || || || || || || || || || || || || || || || || ||

QY  837  TAAATGTATTACAGATGTCTCTGACAAAGATGTCTACTTACTTGAAGCAGCAAAACCGCTG 896
      || || || || || || || || || || || || || || || || || || ||

```

```

||||| 359 TAAGTGCTCACAGACCCCACTAGGTACCTGAGTGCATCAACAGCAACCCGCTG 418
||||| 897 GAACAAGTCCTCTTTAGAGACT 919
||||| 419 GAGCANGCTTACTTTCGGGAAT 441

RESULT 4
AK028582
LOCUS
DEFINITION
Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
library, clone:4732404104 product:similar to DG42III [Brachydanio
rerio], full insert sequence.
ACCESSION
AK028582.1 GI:26324525
VERSION
AK028582.1
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED
10349636
2
AUTHORS
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED
11042159
3
AUTHORS
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kusunagi,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED
11076861
4
AUTHORS
Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Aizawa,K., Iizawa,M., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Iizawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaoka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaudo,I., Pesole,G.,
Quackenbush,J., Schrim,L.M., Staubli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
and Hayashizaki,Y.
TITLE
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
21085660
PUBMED
11217851
5
AUTHORS

```

```

AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
REFERENCE
6 (bases 1 to 4151)
AUTHORS
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,E., Takaku-Akashira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES
Location/Qualifiers
source
1..4151
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM.DB:4732404L04"
/db_xref="taxon:10090"
/clone="4732404L04"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
118..1782
/note="unnamed protein product; putative
similar to DG42III [Brachydanio rerio] (SPTR109DG40,
evidence: FASTA, 75.1%ID, 99.8%length, match=1659)."
/codon_start=1
/db_xref="GI:26324526"
/translacion="MPVQLTALRVVGTSLFALVGLVLAAYVTCYQFIETKHYLS
FGLXGALGLHLLIOSLEAFLEHMRBRRGRPLKHSQSRVACALAAQDEPFL
KCLRSQRATLKNVYVMVDNGRQEDIMDFHEVGLGTQAGFFWNSRNFHAG
EGTEASLQGMERKRVAVVASTFSCIMQKRGKREVMYTAFLAGNSVDYIQVCDSD
TVLPACTEMLEVRLEDPQVGGVDQILNKYDSWISFSSRYWMAFVVERAQS
YFGVQCISGLMYRNSLQQFLEDYHQKFLGSKGSGDGRHLTNRVLSLGRKY
TARSKCTLTPTRYLRNLNQOTWKSYPREMLYNLWLFHKLWMTVSVLGTGPPF
FLATVTLQFGRITWLTLLTQLVGLIKATYACFLGKMGEMFMSLSLXMS
LLPKPATATINKSGWTSGRITVYVNFGLIPVSIWVAVLLGLATYATCQDLFSE
TELAFLVSGRLVGYGVWVALLMLYLALIAIRGCKKPEQYSIAFAY"
4131..4136
/note="putative"
polya_signal
polya_site
983 a 1013 c 1055 g 1100 t
BASE COUNT
ORIGIN
Query Match 3.9%; Score 49.4; DB 11; Length 4151;
Best Local Similarity 49.2%; Pred. No. 2.5;
Matches 159; Conservative 0; Mismatches 161; Indels 3; Gaps 1;

```



RESULT 8  
CNS01MCA  
LOCUS  
CNS01MCA 919 bp DNA linear GSS 14-JUN-2001  
DEFINITION Anopheles gambiae GSS T7 end of clone 21G21 of NotreDame library  
from strain PEST of Anopheles gambiae (African malaria mosquito),  
genomic survey sequence.

REFERENCE  
1 (bases 1 to 919)  
Genoscope.  
AUTHORS  
TITLE  
Direct Submission  
JOURNAL  
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - FRANCE (E-mail : segrefigenoscope.cns.fr  
- Web : www.genoscope.cns.fr)

```

FEATURES             Location/Qualifiers
     source            1..919
                        /organism="Anopheles gambiae"
                        /mol_type="genomic DNA"
                        /strain="PEST"
                        /db_xref="taxon:7165"
                        /clone="21G21"
                        /clone_lib="NotreDame1"
                        /note="end : T7"
BASE COUNT           192 a 165 c 153 g 359 t 50 others
ORIGIN
pasteur.

```

```

/clones="2IG21"
/clone_lib="NotreDame1"
/notes=end : 17"

BASE COUNT      192 a      165 c      153 g      359 t      50 others
ORIGIN

Query Match      3.7%; Score 46.8; DB 29; Length 919;
Best Local Similarity 42.9%; Pred. No. 7.2;
Matches 133; Conservative 26; Mismatches 148; Indels 3; Gaps 1

QY  907  TTCCTTTAGAGATCCCATATTTCTTGTTTAAAGAAATCATGACAAATCCCTTTGTAGCCCTA  966
      ||||| : : : ||| ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db  115  TTTTCTTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT  174

```

[illegible]

Db		235	TTTATTTATTAACTACTTATTTTTTTTTTTTTTTTTW-- -TTTTTCCTWAAATTTTT	291
QY		1087	ATGTTGCCCGTGTCGGACATTCATTACGTCTTAAGCACCCGCTGCCTCTTGTTA	1144
Db		292	ATTTTATTTATTTTWTTTTTTTTACTTATAATATTTTTTACCCTCCTTTTTTATTTTWT	351
QY		1147	TCCTCGTTTTTAGGGGCTGCATTGTTTGCTACAGCCCTGAATATTATATTCCT	1200
Db		352	TTTTATTTTTTACWWTTTTTCCTTTTTTTTTTCYTATATATTTTTTAAATCTTTTTTTTT	411

```

QY      1207 TTTACTATTA 1216
      ||: |||
Db      412 TTWTTACTA 421

RESULT 9
LOCUS   B615669/c
DEFINITION 1702 bp mRNA linear EST 06-FEB-2001
          60234517f1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:445353 5',
          mRNA sequence.
ACCESSION B615669
VERSION   B615669.1 GI:12672372
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1702)
          NIH-MGC http://mhc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-femail.nih.gov
          Tissue Procurement: ATCC
          CDNA Library Preparation: Life Technologies, Inc.
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LILNL at:
          http://image.lilnl.gov
          Plate: L1AM10248 row: m column: 18
          High quality sequence stop: 542.
          Location/Qualifiers
            1..1702
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:445353"
              /tissue_type="hypernephroma, cell line"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_89"
              /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
              Average insert size 1.3 kb. Library enriched for
              full-length clones and constructed by Life Technologies.
              Note: this is a NIH_MGC Library."
BASE COUNT      802 a 286 c 338 g 276 t
ORIGIN
Query Match      3.7%; Score 46.4; DB 10; Length 1702;
Best Local Similarity 48.5%; Pred. No. 9.8;
Matches 128; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY      951 TCTTTTGTAGCCCTATGACCATCTAGGTCGTATGTTATGATGCTGTTATTC 1010
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1585 TCTTTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1526
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1011 TCTGTGGATTCTTTGTAGGCAATGTCAGAGAAATTGATTGGCTCAGGCTTTAGCGCTT 1070
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1525 TCTGTCTCTGTTGTTGTTGTTCTTCTCTGTTATCTCTCTCTCTCTCTCTCTCTCT 1466
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1071 TCTGTGATATCTTCATTTGTCGCGTGTGCGGAACATTCATTACATGCTTAAGCACCC 1130
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1465 CTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1131 GCTGTCTCTCTGTTATCTCCGTTTATGGGGTGTGATGTTGTTCTCTACAGCCCTT 1190
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1405 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1346
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1191 GAAATATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1214
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1345 TCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1322
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
LOCUS   B12074
DEFINITION 1124 bp DNA linear GSS 14-MAY-1997
          F25H9-T7.2 IGF Arabidopsis thaliana genomic clone F25H9, genomic
          survey sequence.
ACCESSION B12074
VERSION   B12074.1 GI:2093194
KEYWORDS  GSS.
SOURCE    Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
          ; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
REFERENCE 1 (bases 1 to 1124)
          Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
          Ecker,J.
          BAC End Sequences at ATGC
          Unpublished
          Other_GSSs: F25H9-T7, F25H9-T7.1, F25H9-Sp6, F25H9-Sp6.1,
          F25H9-Sp6.2
          Contact: Ecker J.
          Arabidopsis Thaliana Genome Center
          University of Pennsylvania
          Dept. of Biology, University of Pennsylvania, Philadelphia, PA
          19104
          Tel: 215-898-9384
          Fax: 215-898-8780
          Email: jecker@genome.bio.upenn.edu
          Seq primer: T7
          Class: BAC ends
          High quality sequence start: 78
          High quality sequence stop: 371.
          Location/Qualifiers
            1..1124
              /organism="Arabidopsis thaliana"
              /mol_type="genomic DNA"
              /strain="Columbia"
              /db_xref="taxon:3702"
              /clone="F25H9"
              /sex="hermaphrodite"
              /clone_lib="IGF"
              /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
              Produced by Thomas Altman"
BASE COUNT      116 a 264 c 61 g 547 t 136 others
ORIGIN
Query Match      3.7%; Score 46.2; DB 28; Length 1124;
Best Local Similarity 33.2%; Pred. No. 10;
Matches 126; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY      837 TAAATGATTTACAGATGTTCTGACAGATGCTACTTACTTGAAGCAGCAAAACCGCTG 896
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      318 TCAAAAGATACAGATGCTGCTGATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNN 377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      897 GAACAAGTCCTTCTTAGAGAGTCCATTATTTCTGTTAAGAAAATCATGAACATCCTTT 956
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      378 NTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 437
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      957 TGTAGCCCTATGGACCATCTAGGTGCTATGTTTATGATGCTGTTATCTGTGCT 1016
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      438 NTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 497
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1017 GGATTTCTTTGTAGGCAATCTCAGACAATTTGATTCGCTCAGGGTTTACGCTTCGCT 1076
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      498 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 557
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1077 GATTATCTTCATGTGTCCTGTGCGAACATTCATTACATGCTTAAGACCGCTGTC 1136
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      558 TNNNTNNNTTTTNTNNNNNTTTTNTNNNTTTTNTNNNTTTTNTNNNTTTNNNNNN 617
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1137 CTCTCTGTTATCTCGGTTTATGGGTGCTGCATTTGTTTGCTACAGCCCTGAAAT 1196
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

	714 bp	mRNA	linear	EST 15-MAY-2003
BX415371.c				
LOCUS				
DEFINITION	BX415371 Homo sapiens THYMUS Homo sapiens cDNA clone CSOCAP005YLI10			
ACCESSION	5-PRIME, mRNA sequence.			
VERSION	BX415371			
KEYWORDS	BX415371.1 GI:30765498			
SOURCE	EST.			
ORGANISM	Homo sapiens (human)			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
JOURNAL	1 (bases 1 to 714)			
COMMENT	Li.W.B., Gruber,C., Jesse,J. and Polayes,D. Full-length cDNA libraries and normalization Unpublished Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cnrs.fr; Web : www.genoscope.cnrs.fr Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : flliang@lifetech.com URL: <a href="http://fulllength.invitrogen.com/">http://fulllength.invitrogen.com/</a> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSOCAP005DP0QPI. Location/Qualifiers 1..714 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /cloned="CSOCAP005YL10" /tissue_type="THYMUS" /clone_lib="Homo sapiens THYMUS" note=Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMWSPORT 6 vector. Library was not normalized." BASE COUNT         303 a   139 c   131 g   97 t   44 others ORIGIN Query Match                  3.6%; Score 45; DB 13; Length 714; Best Local Similarity        40.9%; Pred. No. 17; Matches       99; Conservative   27; Mismatches 116; Indels 0; Gaps 0;  QY       974 TACTTGAGGTGTCATATTGCCTGGTTTAATCTCGTGTGGATTCTCTTAGSCA 1033 Ddb                 ::       ::       ::       ::       ::      709 TAKTKAKGGTGTGTKIXKYGKTGTGGKGKKTTTTTTTTTTTGAGRTTTTTTGTGTTT 650 QY       1034 ATGTCACAGAATAATCATTTGCCAGGGTTTTAGCCCTAAAGCACCCTGCTCACTCACTGAATCTTCATCTGTTG 1093 Ddb       !:: !:: !:: !:: !:: !:: !:: !:: !:: !:: !:: !:: !:: !:: 649 TKGGGKGXYTYTTYTKTKGKANRNTAKNWTWKKBBCBFTTTTGTGTTGTTTITTTT 590 QY       1094 CCGTGTGTCGGAACAATTCATTACATGCTTAAGCACCCGCTGCTCTCTGTTGTTATCTCCGT 1153 Ddb       !: : : : ::  : : : : :   589 KGGTYYVTYTCKTDWTTTTTVVHKWTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 530 QY       1154 TTATGGGCTGCTGCATTTGTTTGCTACAGCCCTTGAAAATATATATCTCTTTTACTA 1213 Ddb         529 TTTTTFTTTTTTTTTTTTTTTTTTTTTYYYCCCCCCCYYTTTTTTTTTTTTTTTTTTT 470 QY       1214 TT 1215 Ddb          469 TT 468  RESULT 13 CNS00CNG/c LOCUS DEFINITION Accession AL059400 Description Drosophila melanogaster genome survey sequence Tet3 end of BAC # BACR26H16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			

```

VERSION      AL059400.1  GI:4946964
KEYWORDS     GSS.
SOURCE       Drosophila melanogaster (fruit fly)
ORGANISM     Drosophila melanogaster
REFERENCE    1 (bases 1 to 939)
AUTHORS      Direct Submission
TITLE        Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
COMMENT      - Web : www.genoscope.cns.fr)
              Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mamoser in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RP1-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA and was constructed by partial
              isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
              PI and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES     Location/Qualifiers
              source
                1..939
                /organism="Drosophila melanogaster"
                /mol_type="genomic DNA"
                /db_xref="taxon:7227"
                /clone="BACR26H16"
                /clone_lib="RP1-98"
                /note="end : TET3"
BASE COUNT   71 a 349 c 104 g 180 t 235 others
ORIGIN
Query Match      3.6%; Score 45; DB 29; Length 939;
Best Local Similarity 6.1%; Pred. No. 18;
Matches 12; Conservative 124; Mismatches 61; Indels 0; Gaps 0;

QY  979  GAGGTCGTCTATGTTATGCTGTTTATTCCTGCTGATGTTCTTTAGGCAATGTC 1038
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db  573  GAKAAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 514
      ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:
QY  1039  AGAGATTGTTGGCTCAGGTTTATGCTTTCCTGCTGATGTTATTCATGTTGTCCTG 1098
      ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:
Db  513  KDDADKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 454
      ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:
QY  1099  TGTGGACATTCATTACATGCTTAAGCACCCGCTGCTCTTCTTCTTATCTCGTTTAT 1158
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db  453  KKAAGAAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 394
      ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:
QY  1159  GGGGTGCTGCTATTTGTT 1175
      ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:  ..:
Db  393  KKKKKKKGGAGDKKK 377

RESULT 14
LOCUS      BX333445/c
DEFINITION BX333445 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC025YN02 5-PRIME, mRNA sequence.
ACCESSION  BX333445
VERSION    BX333445.1  GI:30308208
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1176)

```

```

Li, W. B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9051.r For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DC025DG01QP1
&cluster=9051.r. Contact : peng liang Email : fliang@lifetech.com
URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DC025DG01QP1.
Location/Qualifiers
  1..1176
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="CS0DC025YN02"
  /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
  /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
  /notes="1st strand cDNA was primed with a NotI-oligo(dT)
  primer. Five prime end enriched, double-strand cDNA was
  digested with Not I and cloned into the Not I and EcoR V
  sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT   390 a 176 c 239 g 208 t 163 others
ORIGIN
Query Match      3.6%; Score 44.6; DB 13; Length 1176;
Best Local Similarity 34.3%; Pred. No. 23;
Matches 113; Conservative 51; Mismatches 165; Indels 0; Gaps 0;

QY  900  CAAGTCCTCTTTAGAGAGTCATTATTCCTGTTAAGAAATCATGAACATCCCTTTGT 959
      |||||  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db  476  CAAGGGTCTTTKTCCKCCTTTTYYKWTWTWTCACAAKTATACATAYACCCCTTT 417
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
QY  960  AGCCCTATGACCACTATGAGTGTCTATGTTTATGATGCTGTTTATTCCTGTTGGGA 1019
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db  416  TGGCTTTTTCGACCCCTTGATGCTKTYTCATWYGGGTGTCATGCTGCTGTTTTCG 357
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
QY  1020  TTTCTTTTAGGCAATGTCAGAGAAATTCATGTCCTCAGGCTTTTACGCTTTCTGGTAT 1079
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db  356  TTWCTTTTGPAAITTTTAAATKYVCKAKITTTTYYKKITTTTTHKMATTCCKKTT 257
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
QY  1080  TATCTTCATTTGCGCTGTCGGAACATTCATTACATGCTTAAGCACCCGCTGCTT 1139
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db  296  TTTTTCCTTTTATTTTAAWKCCCKTTCCKCKTTTTCCTTTCCTTCCTGCTGCTGCT 237
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
QY  1140  CTTGCTATCTCCCTTTTATGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db  236  TTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 177
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
QY  1200  TCTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1228
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:
Db  176  CTTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 148
      ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:  ||:

RESULT 15
LOCUS      CNS00100
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR32D23 of RP1-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL068607
VERSION    AL068607.1  GI:4958689
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1101)

```

AUTHORS	TITLE	JOURNAL	COMMENT
Genoscope. Direct Submission	Submitted (02-JUN-1999)	Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : <a href="mailto:secref@genoscope.cns.fr">secref@genoscope.cns.fr</a> ) - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a>	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named PPGI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's FI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .

FEATURES	source
round at http://pacpac.med.buhalo.edu/drosophila_dbac.num.	
Location/Qualifiers	
1. .1101	
/organism="Drosophila melanogaster"	
/mol_type="genomic DNA"	
/db_xref="taxon:7227"	
/clone_lib="BACR32D23"	
/clone_lib="RPC1-98"	
/note_end = TE93"	
BASE COUNT	155 a 166 c 7 g 284 t 489 others
ORIGIN	
Query Match	3.5%; Score 44.4; DB 29; Length 1101;
Best Local Similarity	18.3%; Pred. No. 26;
Matches	73; Conservative 147; Mismatches 178; Indels 0; Gaps 0;
QY	819 TGTTTATCATCACTGCTAAAGTATTACAGATGTCCTGACAAAGATGCTACTTACTT 878
Db	1046 KKKKTTTITKTCTKDWAASAAAWAKTKKTTDTGAKTKTKKGTNNTKTKCKKKTG 987
QY	879 GAAGCAGCAAAACCGCTGGAACAGTCCTCTTTAGAGAGTCATATTCTCTGTTAAGAA 938
Db	986 KRKGDAKAAADDDKDKARAKKGGKKGKWKKGKTKKTKGTTTAKWDD 927
QY	939 AATCATGACAATTCCTTTTGTAGCCCTATGAGACATACTTTGAGGTGCTATGTTTATGAT 998
Db	926 DAWMDWADTDWKGAGGGGGKGDWTDTDWCKDAWKKGDKDTKKKKKGGGKAKGAA 867
QY	999 GCTTGTTTATCTGTGTGGATTTCTTTGTAGCAATGTCAGAGAAATTGATGGGTCAAG 1058
Db	866 AAADRKGAADGAAAKTKTKKAKTWKKKGGGWRGKGTGKGGKGGGTAKGUTKK 807
QY	1059 GGTTTAGCCTTCTGTGGTATATCTTCATGTTGTGCCCTGTGTCGGAACATTCATACAT 1118
Db	806 KGRKKKGTGDKRTWTKTTTGTGTGTAWTKTGTGKKKKKGGKGGKGDWTKGTWDT 747
QY	1119 GGTTAAGCACCCGCTGCTCCTCTGTTATCTCCGTTTATGGSGTGTGCGATTTGTTCT 1178
Db	746 KTKTKGKTGGGGKKTKYKGYKTNWAAWAKAKTKTKTKGKTGAKTADTKKKKKKD 687
QY	1179 CCTACAGCCCTTGAATATATATCTCTTTTACTATTA 1216
Db	686 KGTGKTGKKTKGAAWNGTGDFTKDKKKKAAWND 649

Search completed: October 2, 2003, 11:19:44  
Job time : 2975 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 2, 2003, 10:30:08 ; Search time 361 Seconds  
(without alignments)  
3118.187 Million cell updates/sec

Title: US-09-469-200d-2  
Perfect score: 2145  
Sequence: 1 MRLTKNLITVAFSIFWVLL.....KLYSLFIRNADWGRKKLL 417

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPO.spool/US09469200/runat\_02102003\_090347\_22770/app-query.fasta.1.583  
-DB=N\_Geneseq\_19Jun03 -QPM=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosumb2 -TRANS=human40.cdi  
-LIST=45 -LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09469200.@CGN.1.1.312 -runat\_02102003\_090347\_22770 -NCPU=6 -ICPU=3  
-NO\_MMAPP -LARGEQUERY -NEG\_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2145	100.0	1254	20	AAX58841 Streptococcus equi
2	1586.5	74.0	1335	24	ABN67167 Streptococcus poly
3	1586.5	74.0	1512	25	ABN57951 S. pyogenes hyalur
4	481.5	22.4	7193	21	AAA14914 DNA encoding nodul
c 5	463	21.6	534720	19	AAV30458 Rhizobium species
c 6	463	21.6	536165	19	AAV30458 Rhizobium species
7	447.5	20.9	1662	25	ACC51023 Human bladder canc
8	447.5	20.9	1662	25	ACC51023 Lung cancer-associ
9	447.5	20.9	1662	25	ABX76214 Lung cancer-associ
10	440	20.5	1665	19	AAV18822 Murine hyaluronan
11	440	20.5	1665	21	AAZ88201 Mouse hyaluronate
12	440	20.5	1665	25	ABZ76736 Mouse hyaluronan s
c 13	439	20.5	1653	24	ABL41013 Murine hyaluronan
14	439	20.5	1659	25	ABZ76735 Mouse hyaluronan s
15	439	20.5	2947	19	AAV18821 Murine hyaluronan
16	439	20.5	2948	24	AB199697 Mouse ischaemic co
17	433.5	20.2	3003	24	ABK84373 Human cDNA differe
18	433.5	20.2	4018	22	AAU26639 Human breast cance
19	428.5	20.0	2890	21	AAZ88200 Mouse hyaluronate
20	370.5	17.3	2117	18	AAZ96713 Human hyaluronan s
21	367	17.1	1752	25	ABZ76734 Mouse hyaluronate
22	367	17.1	2102	18	AAZ91655 Murine hyaluronan s
23	367	17.1	2102	20	AAZ10862 Hyaluronate syntha
24	367	17.1	2102	21	AAA39987 Murine HAS1 cDNA
25	367	17.1	2102	21	AAZ88199 Mouse hyaluronate s
26	366.5	17.1	2116	18	AAZ99541 Human hyaluronan s
27	336.5	15.7	1740	20	AAX58846 Chlorella virus PB
c 28	265.5	12.4	265118	22	AAH41227 Pyrococcus abyssi
c 29	265.5	12.4	349980	22	AAH41226 Pyrococcus abyssi
30	262.5	12.2	2365589	24	ABA90521 Genomic sequence o
31	245.5	11.4	1239	22	AAC89348 Staphylococcus aur
32	245.5	11.4	1239	22	AAC89349 Staphylococcus aur
33	245.5	11.4	5648	22	AAC89353 Staphylococcus aur
34	242.5	11.3	4951	17	AAU05848 Polycistronic cps
35	242.5	11.3	4951	20	AAZ11802 Nucleotide sequenc
c 36	242.5	11.3	7430	18	AAV74571 Staphylococcus aur
37	234.5	10.9	1257	24	ABN91751 Staphylococcus epi
38	234.5	10.9	4500	21	AAZ87998 Nucleotide sequenc
39	233.5	10.9	570	24	ABN96952 Gene #3450 used to
40	226	10.5	1266	24	ABN67166 Streptococcus poly
41	226	10.5	1266	24	ABN70371 Streptococcus poly
c 42	226	10.5	215561	24	ABN71527 Streptococcus poly
43	198.5	9.3	1203	24	ABN71063 Streptococcus poly
c 44	195.5	9.1	13884	20	AAX13278 Enterococcus faeca
c 45	195.5	9.1	13884	24	ABN99073 Enterococcus faeca

# ALIGNMENTS

RESULT 1  
AAX58841  
ID AAX58841 standard; DNA; 1254 BP.  
AC AAX58841;  
XX  
XX  
XX 16-AUG-1999 (first entry)  
DT  
DE Streptococcus equisimilis hyaluronate synthase seHAS DNA.  
XX  
XX Hyaluronate synthase; seHAS; hyaluronic acid; hyaluronan; ss.  
XX  
XX Streptococcus equisimilis.  
XX  
XX Key Location/Qualifiers  
FT primer\_bind complement (316..337)  
FT /\*tag= a







Claim 1; Fig 1; 47pp; English.

CC The present sequence encodes nodulation efficiency factors. The present  
 CC sequence is a 7.2 kb EcoRI/BamHI fragment from *Sinorhizobium meliloti*  
 CC USDA 1170, that gives S. meliloti strain NRG 185 the ability to nodulate  
 CC at least 50% of inoculated *Medicago lacinata* (cut-leaf medic) plants  
 CC within 10 days of inoculation. The fragment contains nod genes nodA,  
 CC nodB, nodC, nodD, nodJ, nodJ. The nodulation factors are used to  
 CC increase the nodulation efficiency of *Sinorhizobium* for *Medicago*  
 CC *lacinata*. The nodulation factors are also used to improve nitrogen  
 CC fixation in legumes.  
 XX

SQ Sequence 7193 BP; 1434 A; 2094 C; 2132 G; 1532 T; 1 other;

## Alignment Scores:

Pred. No.: 2,47e-39 Length: 7193  
 Score: 481.50 Matches: 133  
 Percent Similarity: 49.50% Conservative: 65  
 Best Local Similarity: 33.25% Mismatches: 169  
 Query Match: 22.45% Indels: 33  
 DB: 21 Gaps: 10

US-09-469-200D-2 (1-417) x AAAL4914 (1-7193)

QY 33 SerLeuSerIleTyrGlyPheLeuLeuIleAlaTyrLeuLeuValLysMetSerLeuSer 52  
 DB 1811 GCATCTCGATTACGGCGTCTCTTGCCGCGCTAC-----AGGACATGCAA 1858  
 QY 53 PhePheTyr-----LysProPheLysGlyArg 61  
 DB 1859 GTCTATATACTCGCGCGATAAACGGTCCACCGCTGTCGGCAGAACCGATCGAGACCGC 1918  
 QY 62 AlaGlyGlnTyrLysValAlaAlaIlePheProSerTyrAsnGluAspAlaGluSerLeu 81  
 DB 1919 GCCCTCTGCGCCGCTGGACGTTATCTGTCGCCAGTTTCAATCAGGACCCAGCATGCTC 1978  
 QY 82 LeuGluThrLeuLysSerValGlnGlnThrTyrPro---LeuAlaGluIleTyrVal 100  
 DB 1979 TCGCGCTCGCTCGGCTCCATTCACACACGAGATTCCTGGAGACTGCGAGTCTATGTC 2038  
 QY 101 ValAspAspGlySerAlaAspGluThrGlyIleLysArgIleGluAsp---TyrValArg 119  
 DB 2039 GTTGATGATGCTCTCGGAACCGCGAGCAATCGTCGTTGACAGATTCTTATTCGCGC 2098  
 QY 120 AspThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLys 139  
 DB 2099 GAT-----CCGAGGTTACCTTCATCTCTCCCGAGAGACGTCGGAAG 2143  
 QY 140 ArgHisAlaGlnAlaThrAlaPheGluArgSerAspAlaAspValPheLeuThrValAsp 159  
 DB 2144 CGGAAGCGCAGATTGCCGATAGTCAATCTCTGGGAACTGGTGTGAATGTCGAC 2203  
 QY 160 SerAspThrTyrIleTyrProAspAlaLeuGluLeuLysThrPheAsnAspPro 179  
 DB 2204 TCAGACAGCAGANTGCTTTCGATGCTGCTCCAAAGCTTGCTCGAAGATCGCAATCCA 2263  
 QY 180 ThrValPheAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThr 199  
 DB 2264 GAGGTGTGGCGCCATGGTCACTACAGCTAGCAATCGGAGTGACACTGGCTGACG 2323  
 QY 200 ArgLeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaGlnSerVal 219  
 DB 2324 AGATTGATCGACATGGATGTTGCTTGCCTGCAAGAGAGCGCGCGCACAGGCTCGC 2383  
 QY 220 ThrGlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValValVal 239  
 DB 2384 TTCGGTGGCGTTATGTGTGCTGCGGCCCATGTGCTATGCTACCGTCTCGGCTCGCT 2443  
 QY 240 ProAsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAsp 259  
 DB 2444 GCGCTGCTGACAGTACGAACGCGCGTGTTCGCGGTAAACCAACGCACTTCGGGTGAG 2503  
 QY 260 AspArgCysLeuThrAsnTyrAlaThrAspLeuGly---LysThrValTyrGlnSerThr 278  
 DB 2504 GATCGTCATCTGAGGATCTCATGTTAGGACGAGGTTTAGACCGAGATGTTCCAGAC 2563

QY 279 AlaLysCysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArg 298  
 DB 2564 GCCATAGTGGCAGCGTGTTCGCGATAGGCTGAAGCGGTATCTAGCCCAACAACGCTG 2623  
 QY 299 TrpAsnLysSerPheArgGluSerIleIleSerValLysLysIleMetAsn---Asn 317  
 DB 2624 TGGCAGCAGACGCTTCGCGGACAGCTTTTGGCGTCCCTCTGTGGGCGGCGCTCAAC 2683  
 QY 318 ProPheValAlaLeuTyrThrIleLeuGluValSerMetPheMetLeuValTyrSer 337  
 DB 2684 GCTTATCTCATCTCGATGCGGTGGCGCAGACTATCGGCCCATTTGCTGCCCTCTCG 2743  
 QY 338 ValValAspPhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPhe 357  
 DB 2744 ATAGTGACG-----GGACTTGGCATTTTCATCATGACCGCCACACATCGCGTGG 2791  
 QY 358 LeuValIleIlePheIleValAlaLeu-----CysArgAsnIleHisTyrMet 373  
 DB 2792 TGGCAATTTTGTGATGTCATCCATGCGCCATGATACGTCGACCGTGGTGGCATGAAT 2851  
 QY 374 LeuLysHisProLeuSerPheLeuLeuSerProPheTyrGlyValLeuHisLeuPheVal 393  
 DB 2852 GCTCGCCAA---CTTAGATTCTTGGTACGTTCTACACACACCCCATCAACCTCTTCTC 2908  
 QY 394 LeuGluProLeuLysLeuTyrSerLeuPheThrIleArgAsnAlaAspTrpGlyThrArg 413  
 DB 2909 CTAATTCGTTGAAAGCCTATGCGTTGTATCATTTGTCAATGTCACAGGAGCTGTTGTCACGC 2968  
 RESULT 5  
 AAV30458/C  
 ID AAV30458 standard; DNA; 534720 BP.  
 XX AAV30458;  
 AC  
 XX  
 DT 14-OCT-1998 (first entry)  
 XX  
 DE Rhizobium species plasmid pNGR234a.  
 XX  
 KW Symbiosis; open reading frame; ORF; plasmid; vector; transportation;  
 KW degradation; metabolism; host range; nitrogen fixation; nodulation;  
 KW legume; plant; ds.  
 XX  
 OS Rhizobium sp.  
 XX  
 FH Location/Qualifiers  
 FT CDS 417796..418671  
 FT /tag= a  
 FT /standard\_name= "ORF K1"  
 FT /product= "oligopeptide permease"  
 FT /note= "homologous to the oppC gene"  
 FT CDS 418673..419680  
 FT /tag= b  
 FT /standard\_name= "ORF K2"  
 FT /product= "oligopeptide permease"  
 FT /note= "homologous to the oppD gene"  
 FT CDS 419677..420738  
 FT /tag= c  
 FT /standard\_name= "ORF K3"  
 FT /product= "oligopeptide permease"  
 FT /note= "homologous to the oppF gene"  
 FT CDS 420774..422159  
 FT /tag= d  
 FT /standard\_name= "ORF K4"  
 FT /product= "encapsulation-like protein"  
 FT /note= "homologous to the CapA gene"  
 FT CDS 422628..424031  
 FT /tag= e  
 FT /standard\_name= "ORF K5"  
 FT /product= "aminotransferase-like protein"  
 FT /note= "homologous to the BioA gene"  
 FT CDS 424056..425594  
 FT /tag= f



QY 258 GlyAspAspArgCysLeuThrAsnTyrAlaThrAspLeuGly---LysThrValTyrGln 276  
|||:::||||| ||||| :||| :||| :||| :|||

Db 156562 CCGGAGCCATCGGCGGACGGTCTCCAAACTCGATGGGCGCTATCTGGCCACAA 156503

The present invention describes a method for detecting a bladder cancer-associated transcript in a cell from a patient. The method comprises contacting a biological sample from the patient with a polynucleotide that selectively hybridizes to a sequence that is 80 % identical to a table of sequences (see ACC530951 to ACC531059). ACC530951 to ACC531059 encode the human bladder cancer-associated proteins given in AB481416 to AB482422. Bladder cancer-associated sequences from the present invention have cytostatic activities, and can be used in antisense gene therapy and in vaccine production. The method can be used for detecting a bladder

QY		4	LeuLysAsnLeuIleThrValValAlaPheSerIlePhe-----Trrp	17
Db		13	CTGACGACAGCCCTGTGGTGTCGGCGCACCAAGCCTGTTTGCCCTGGCAGTGCATGCTGGGTGGC	72
QY		18	ValLeuLeuIleThrValAsnValTyrrLeuPhe-----GlyAlaLysGlySerLeuSer	35
Db		73	ATCCTGGCAGCCTATTGTGACGGGTACCAGTGTTCATCCACACGGAAAGACACTACTCTGCC	132
QY		36	-----IleTyrrGlyPheLeuLeuIleAlaTyrrLeuValLysMetSerLeuSerPhe	53
Db		133	TTCGGCCTGTACGGGCCCATCTCGGGCTGCACCTGCTCATTCAGAGCCCTTTTGCCTTC	192
QY		54	PheTyrrLysProPheLysGlyArgAlaGlyGlnTyrr-----	65
Db		193	CTGGAGCACCGGCGCATGTGGAGTGGCGGCGACGCCCTGAAGCTGCCCTCCCOCGGCGCG	252
QY		66	---LysValAlaAlaIleIleProSerTyrrAsnGluAspAlaGluSerLeuLeuThr	84
Db		253	GCGTCGGTGGCAGTGTGCATTCGCCGTACCGAGGAGCCCTGACTACTTGGCGAAGTGC	312
QY		85	LeuLysSerValGlnGlnThrTyrrProLeuAlaGluIleTyrrValValAspAspGly	104
Db		313	CTGCGCTCGGCCACGAGCGCATCTCCTCTGACCTCAAGTGGTCAATGTTGGTGGTAGTGC	372
QY		105	SerAlaAspGlu-----	108
Db		373	AACGCCAGGAGGACGCGCTACATCTGTGGACATCTCCACGAGGTGCTGGCGGCGCACGAG	432
QY		108	-----	108
Db		433	CAGCGCGGCTCTTTGTGTGGCGCAGCAACTTCCATGAGCGACGCCAGGGTGACACGGAG	492
QY		109	-----ThrrGlyLysArgIleGluAspTyrrValArgAspThrGlyAspLeu	124
Db		493	GCCAGCCTGCAGGAGGSCATGACCGTGTCCGGATGTGTGCGG-----	537
QY		125	SerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArgHisAlaAla	144
Db		538	---GCCACACCTTCGTGTCATCATCAGAGTGGGAGCGAACGCGCGAGTGCATGTAC	594
QY		145	TrpAlaPheGluArg-----SerAspAlaAspValPheLeuthrValAspSerAspThr	162
Db		595	ACGGCTTCACAGGCCCTCGCGCGATTCGGTGGACTACATCCAGGTGTGGCACTCTGCACACT	654
QY		163	TyrrIleTyrrProAspAlaLeuGluIleLeuLysThrPhe---AsnAspProThrVal	181
Db		655	GTGCTGGATCCAGCCCTGCACCATCGAGATCCTCGAGTCTCGAGGAGGAGTCCCCCAAGTA	714
QY		182	PheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuthrArgLeu	201
Db		715	GGGGAGTCGGGGAGATGTCCAGATCCTCAACAGTACGACTCATGTGGATTCCTTCTCTG	774
QY		202	ThrAspIleArgTyrrAspAsnAlaPheGlyValGluArgAlaLaGlnSerValThrGly	221

PR 13-NOV-2001; 2001US-350666P.  
 PR 29-NOV-2001; 2001US-334370P.  
 PR 12-APR-2002; 2002US-372246P.  
 XX (EOSB-) EOS BIOTECHNOLOGY INC.

PI Aziz N, Murray R;

DR WPI; 2003-093161/08.

DR P-PSDB; AB056686.

XX Detecting a lung cancer-associated transcript in a cell from a patient  
 PT for treating lung cancer, by contacting a biological sample from the  
 PT patient with a polynucleotide that exhibits increased or decreased  
 PT expression in lung cancer.

XX Claim 22; Page 401-402; 453pp; English.

XX The invention relates to a method for detecting a lung cancer-associated  
 CC transcript in a cell from a patient, comprising contacting a biological  
 CC sample from the patient with a polynucleotide that selectively hybridizes  
 CC to a sequence that is at least 80 % identical to a gene that exhibits  
 CC increased or decreased expression in lung cancer samples. Lung  
 CC cancer-associated polynucleotides and polypeptides are used for  
 CC identifying a compound that modulates a lung cancer-associated  
 CC polypeptide, for inhibiting proliferation of a lung cancer-associated  
 CC cell to treat lung cancer in a patient and for treating a mammal having  
 CC lung cancer by administering a modulatory compound identified. The  
 CC methods are useful for treating lung cancer, such as small cell lung  
 CC cancer, non-small cell lung cancer or other benign or precancerous  
 CC lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive  
 CC pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial  
 CC pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides  
 CC and polypeptides are useful for diagnostic purposes and as targets for  
 CC screening for therapeutic compounds that modulate lung cancer, such as  
 CC antibodies. Sequences ABX76124-ABX76474 represent lung cancer-associated  
 CC polynucleotides of the invention.

XX SQ Sequence 1662 BP; 300 A; 501 C; 486 G; 375 T; 0 other;

#### Alignment Scores:

Pred. No.: 1-12e-36 Length: 1662  
 Score: 447.50 Matches: 134  
 Percent Similarity: 44.33% Conservative: 85  
 Best Local Similarity: 27.13% Mismatches: 172  
 Query Match: 20.86% Indels: 103  
 DB: 25 Gaps: 16

US-09-469-200D-2 (1-417) x ABX76415 (1-1662)

QY 4 LeuLysAsnLeuIleThrValValAlaPheSerIlePhe-----Trp 17  
 DB 13 CTGACGACAGCCCTGCGTGTGGTGCCACAGCCTGTTGCCCTGGCAGTGTGGTGC 72  
 QY 18 ValLeuIleThrValAsnValTyLeuPhe-----GlyAlaLysGlySerLeuSer 35  
 DB 73 ATCCCTGGCAGCCTATGTCAGGGCTACCAAGTTATCCACCGGAAAGCACTACCTGTCC 132  
 QY 36 -----IleTyGlyPheLeuLeuIleAlaTyLeuValLysMetSerLeuSerPhe 53  
 DB 133 TTGGCGCTGTACGGCGCCATCCCTGGGCTGCACCTGCTATTACAGAGCCTTTTGGCTTC 192  
 QY 54 PheTyLysProPheLysGlyArgAlaGlyGlnTyTyr----- 65  
 DB 193 CTGGAGCACCGGCGCATGCGAGCTGCGGGCCAGCCCTGAAGCTGCCCTCCCGGGCGG 252  
 QY 66 ---LysValAlaAlaIleIleProSerTyRasnGluAspAlaGluSerLeuLeuThr 84  
 DB 253 GGCTCGGTGGCACTGTGCATTGCCGCTACCGAGGAGCCCTGACTACTTGCACAAGTGC 312  
 QY 85 LeuLysSerValGlnGlnThrTyRProLeuAlaGluIleTyRValValAspAspGly 104  
 DB 313 CTCGCGTGGCGGACGCGATCTCTCCCTGACCTCAAGGTGGTCATGGTGGTATGCG 372

QY 105 SerAlaAspGlu----- 108  
 DB 373 AACCGCCAGGAGGAGCGCTACATGCTGGACATCTTCCACGAGGTGCTGGGGGACCGAG 432  
 QY 108 ----- 108  
 DB 433 CAGCGCGGCTTCTTGTGTGGCGCAGCAACTTCCATGAGGCGGCGGCTGAGACGGAG 492  
 QY 109 -----ThrGlyIleLysArgIleGluAspTyRValArgAspThrGlyAspLeu 124  
 DB 493 GCCAGCCTGCGAGGAGGATGACCGTGTGGGATGTGTGGCG----- 537  
 QY 125 SerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArgHisAlaGlnAla 144  
 DB 538 ---GCCAGCACCTTCTGTCATCATGCAAGTGGGGAGGCAAGCGGCGGTGATGATC 594  
 QY 145 TrpAlaPheGluArg-----SerAspAlaAspValPheLeuThrValAspSerAspThr 162  
 DB 595 ACGGCTTCAAGGCGCTCGGGATTCGGTGGACTACATCCAGGTGCGGACTCTGACACT 654  
 QY 163 TyrIleTyRProAspAlaLeuGluGluLeuLeuLysThrPhe---AsnAspProThrVal 181  
 DB 655 GTGTGTGATCCAGCCTGCGACCATCGAGATGCTGAGTCTCGAGGAGGATCCCAAGTA 714  
 QY 182 PheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeu 201  
 DB 715 GGGGAGTGGGGAGATGTCAGATCTCCACAAAGTACGACTCATGGATTCTCTCTG 774  
 QY 202 ThrAspIleArgTyRAspAlaPheGlyValGluArgAlaAlaGlnSerValThrGly 221  
 DB 775 AGCAGCTGCGGTACTGTGATGGCTTCAACGTGGAGGGGCTGCCAGTCTACTTTGGC 834  
 QY 222 AsnIleLeuValCysSerGlyProLeuSerValTyRArgGluValValProAsn 241  
 DB 835 TGTGTGCGAGTGATAGTGGCCCTTGGCATGTACCGCAACAGCCTCTCCACAGTTC 894  
 QY 242 IleAspArgTyRLeuAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArg 261  
 DB 895 CTGGAGGAGTGTGATCCATCAGAGTCTTAGGCGACAGTGCAGCTTGGGGATGCCGG 954  
 QY 262 CysLeuThrAsnTyRAlaThrAspLeuGly---LysThrValTyRGlinsThrAlaLys 280  
 DB 955 CACCTCACCAACCGAGTCTCTGAGCTGGCTACCGAACTAAGTATACCGCGGCTCCAA 1014  
 QY 281 CysIleThrAspValProAspLysMetSerThrTyRLeuLysGlnGlnAsnArgTyRAsn 300  
 DB 1015 TGCCTCACAGAGACCCCACTAAGTACCTCGGTGGCTCAACACGAAACCGCGTGGAGC 1074  
 QY 301 LysSerPheArgGluSerIle----- 308  
 DB 1075 AAGTCTTACTCGGGAGTGGCTCTACAACCTCTCTGTGGTTCCTTAAGCACCACTCTGG 1134  
 QY 309 IleSerValLysIleMetAsnAsn-----ProPheValAlaLeuTyRThrIleLeu 326  
 DB 1135 ATGACCTACAGACAGTGGTTCACGGGTTCCTCCCTCTCTCTCTTCCATGCCAGGTTATA 1194  
 QY 327 GluValSerMetPheMetMetLeuValTyRValValAspPhePheValGlyAsnVal 346  
 DB 1195 CAGCTT-----TTCACGGGGCGGCATC 1218  
 QY 347 ArgGluPheAspTyRLeuArgValLeuAlaPheLeuValIleIlePheIleValAlaLeu 366  
 DB 1219 -----TGG---AACATTCTCTCTCTGCTGACGGTGCAGCTGGTGGGCTATT 1263  
 QY 367 CysArgAsnIleHis---TyrMetLeuLysHisProLeuSerPheLeuLeuSerProPhe 385  
 DB 1264 ATCAAGCCACCTACCTGCTTCTTCGGGCAATGCAGAGATGATCTTCATGCTCCCTC 1323  
 QY 386 TyRGlYValLeuHisLeuPheValGlnProLeuLysLeuTyRSerLeuPheThrIle 405  
 DB 1324 TACTCTCTCTATATGTCAGCCTTCTGCGGCGCAAGATCTTTGCCATTGCTACCATC 1383

QY 406 ArgAsnAlaAspTTPGlyThr-----ArgLysLysLeuLeu 417  
 Db 1384 AACAAATCTGGCTGGCGACCTCTGGCGGAAAMACCATGTG 1425

RESULT 9

ID ABX76214 standard; DNA; 4049 BP.

XX AC ABX76214;

DT 02-APR-2003 (first entry)

XX Lung cancer-associated polynucleotide #83.

Lung cancer-associated polynucleotide: gene; ds; cytostatic; emphysema;  
 antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;  
 small cell lung cancer; benign lesion; precancerous lesion; bronchitis;  
 chronic obstructive pulmonary disease; hypersensitivity pneumonitis;  
 interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.

XX Unidentified.

XX WO200286443-A2.

XX 31-OCT-2002.

XX 18-APR-2002; 2002WO-US12476.

XX 18-APR-2001; 2001US-284770P.

XX 10-MAY-2001; 2001US-290492P.

XX 09-NOV-2001; 2001US-339245P.

XX 13-NOV-2001; 2001US-350666P.

XX 29-NOV-2001; 2001US-334370P.

XX 12-APR-2002; 2002US-372246P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Aziz N, Murray R;

XX WPI; 2003-093161/08.

XX P-PSDB; ABU56490.

PT Detecting a lung cancer-associated transcript in a cell from a patient  
 for treating lung cancer, by contacting a biological sample from the  
 patient with a polynucleotide that exhibits increased or decreased  
 expression in lung cancer

PS Claim 22; Page 252-253; 453pp; English.

XX The invention relates to a method for detecting a lung cancer-associated  
 transcript in a cell from a patient, comprising contacting a biological  
 sample from the patient with a polynucleotide that selectively hybridizes  
 to a sequence that is at least 80 % identical to a gene that exhibits  
 increased or decreased expression in lung cancer samples. Lung  
 cancer-associated polynucleotides and polypeptides are used for  
 identifying a compound that modulates a lung cancer-associated  
 polypeptide, for inhibiting proliferation of a lung cancer-associated  
 cell to treat lung cancer in a patient and for treating a mammal having  
 lung cancer by administering a modulatory compound identified. The  
 methods are useful for treating lung cancer, such as small cell lung  
 cancer, non-small cell lung cancer or other benign or precancerous  
 lesions, e.g. atelectasis, emphysema, bronchitis, chronic obstructive  
 pulmonary disease, fibrosis, hypersensitivity pneumonitis, interstitial  
 pulmonary fibrosis, asthma and bronchiectasis. The genes, polynucleotides  
 and polypeptides are useful for diagnostic purposes and as targets for  
 screening for therapeutic compounds that modulate lung cancer, such as  
 antibodies. Sequences ABX76214-ABX76474 represent lung cancer-associated  
 polynucleotides of the invention.

XX Sequence 4049 BP; 919 A; 1046 C; 1049 G; 1035 T; 0 other;

Alignment Scores:

Pred. No.: 3.92e-36 Length: 4049

Score:	447.50	Matches:	134
Percent Similarity:	44.33%	Conservative:	85
Best Local Similarity:	27.13%	Mismatches:	172
Query Match:	20.86%	Indels:	103
DB:	25	Gaps:	16
US-09-469-200D-2 (1-417) x ABX76214 (1-4049)			
QY	4	LeuLysAsnLeuLeuThrValValAlaPheSerIlePhe-----	-----TriP 17
Db	13	CTGACGACAGCCCTGGTGGTGGGACACGCTGTTGGCTGGCAGTGTGGGTGGC	72
QY	18	ValLeuLeuIleLeuValAsnValTyrLeuPhe-----GlyAlaLysGlySerLeuSer	35
Db	73	ATCTCTGGACCTATGTGTACGGGCTACCACTTCCACACGAGAAAGCACTA	132
QY	36	-----IleTyrGlyPheLeuLeuAlaTyrLeuLeuValLysMetSerLeuSerPhe	53
Db	133	TTCCGGCTGTACGGCGCCATCTCTGGCCCTGACCTGTCTATTCAGAGCTTTTGGCTTC	192
QY	54	PheTyrLysProPheLysGlyArgAlaGlyGlnTyr-----	65
Db	193	CTGGAGCACGGCGCATCGCACGTGCGCGCCAGCCCTGAAGCTGCCCTCCCGCGCG	252
QY	66	---LysValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSerLeuLeuThr	84
Db	253	GGCTCGGTGGCACTGTGCATTCGCCATACACAGAGGAGCCCTGACTACTTCCGAAGTGC	312
QY	85	LeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrValValAspGly	104
Db	313	CTGGCTCGGCCACGCGCATCTCTCCCTGACCTCAAGTGTGTGTGGTGGTGGC	372
QY	105	SerAlaAspGlu-----	108
Db	373	AACGCCAGGAGGACGCGCTACATCTTCCACGAGGTGTGGCGGCGCACCGAG	432
QY	108	-----	108
Db	433	CAGCGCGGCTCTTTGTGGCGCACACTTCCATGAGCGAGGCGAGGTGAGACGGAG	492
QY	109	-----ThrGlyIleLysArgIleGluAspTyrValArgAspThrGlyAspLeu	124
Db	493	GCCAGCTCGCAGGAGGCGATGGCGGTGTGGGATGTGTGGCGG-----	537
QY	125	SerSerAsnValIleValHisArgSerGlnLysAsnGlnLysArgHisAlaGlnAla	144
Db	538	---GCCAGCACCTCTCTGTGCATCATGCGAAGTGGGGAGGCAAGCGCGGTCATGTAC	594
QY	145	TrpAlaPheGluArg-----SerAspAlaAspValPheLeuThrValAspSerAspThr	162
Db	595	ACGGCTCTCAAGCCCTCGCGGATTCGGTGACTACATCCAGGTGTGGCATCTTGACAT	654
QY	163	TyrIleTyrProAspAlaLeuGluLeuLeuLysThrPhe---AsnAspProThrVal	181
Db	655	GTGCTGGATCCACGCTGCACCATCGAGATGCTCGAGTCTCTGGAGGAGATCCCCAAGTA	714
QY	192	PheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeu	201
Db	715	GGGGAGTGGGGAGATGTCCAGATCTCTCAACAAAGTACGACTCATGTGATTTCTCTCTG	774
QY	202	ThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThrGly	221
Db	775	AGCAGCGTGGGTACTGATGGCTTCACTGAGGGGCGCTGCCAGTCTCTACTTTGGC	834
QY	222	AsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValValProAsn	241
Db	835	TGTGTGACAGTGTATTAGTGGCCCTTGGCATGTACCGCAACAGCCCTCCCTCCAGAGTTC	894
QY	242	IleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspArg	261
Db	895	CTGGAGGACTGGTACATCAGATCTCTAGGACGACAGTGCAGCTTCGGGGATGACCGG	954
QY	262	CysLeuThrAsnTyrAlaThrAspLeuGly---LysThrValTyrGlnSerThrAlaLys	280



Db 655 ACTGTCGGACCCAGCCGTCACCATTCAGATCTTCAGCTCTGGAGAGATCCCAA 714  
 QY 181 ValPheAlaIaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200  
 Db 715 GTAGAGGTGTGGAGAGATGCCAAATCCTCAACAGATGATTCATGGATCTCCTTC 774  
 QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThr 220  
 Db 775 CTGACGAGTGTGAGTACTGGATGGCTTCAAGTGGAGCGGCTCCAGTCCACTTT 834  
 QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValValPro 240  
 Db 835 GGCTGTGTGAATGATTAGTGGCTTTGGCATGTACCGCAACAGCCTCCTTCAGAG 894  
 QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260  
 Db 895 TTCCCTGGAGATGGTACCATCAGAACTCTTAGCGCAAGTGCAGCTTTGGGATGAT 954  
 QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGly---LysThrValTyrGlnSerThrAla 279  
 Db 955 CGGCACCTTACCAACCGATCTGCTGCTGGCTACCGCACTAAGTATACAGCGCTCT 1014  
 QY 280 LysCysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTyr 299  
 Db 1015 AAGTGCCTCAGAGACCCCACTAGGTACCTTCGATGGCTCAATCAGCAACCCGCTGG 1074  
 QY 300 AsnLysSerPhePheArgGluSerIle----- 308  
 Db 1075 ACCAGTCTTACTTTCGGGATGCGCTACAAATCTCTGTGGTTCCTCAAGCACCACTC 1134  
 QY 309 ---IleSerValLysLysIleMetAsnAsn-----ProPheValAlaLeuThrPheIle 325  
 Db 1135 TGGATGACCTATGAATCAGTGTGCACAGCTTCTCCCATCTTCTCTCATCTACATC 1194  
 QY 326 LeuGluValSerMetMetMetLeuValTyrSerValAspPheValGlyAsn 345  
 Db 1195 ATACAACTT-----TTCTACCGTGGCGC 1218  
 QY 346 ValArgGluPheAspTyrLeuArgValLeuAlaPheLeuValIleIlePheIleValAla 365  
 Db 1219 ATC-----TGG---AACATTCCTCTCTCTCTACAGTGCAGTGTGGGC 1263  
 QY 366 LeuCysArgAsnIleHis---TyrMetLeuLysHisProLeuSerPheLeuLeuSerPro 384  
 Db 1264 ATTATCAAGCTACCTATGCTGCTCTCTCGAGGCAATGCAGAGATGATCTTCATGTC 1323  
 QY 385 PheTyrGlyValLeuHisPheValLeuPheValLeuAlaPheLeuValIleIlePheIleValAla 365  
 Db 1324 CTTACTCCCTTCTCTATATGCTCAGGCTCTTGCAGCAAGATCTTTGCTATTGCTACC 1383  
 QY 405 IleArgAsnAlaAspTyrPheIleThr-----ArgLysLysLeuLeu 417  
 Db 1384 ATCAACAAGTCTGGCTGGGCACTCTCTGGCAGGAAACCAATGTC 1428  
 RESULT 11  
 ID AAZ88201 standard; cDNA to mRNA; 1665 BP.  
 XX AC AAZ88201;  
 XX AC AAZ88201;  
 DT 27-APR-2000 (first entry)  
 DE Mouse hyaluronate synthase modified protein HAS3 encoding cDNA.  
 XX Mouse; murine; hyaluronate synthase; modification; HAS1; HAS2; HAS3;  
 KW research reagent; biochemical research; medical development; ds.  
 XX OS Mus sp.  
 XX Key Location/Qualifiers  
 FT 1..1665  
 FT CDS /\*tag= a

FT  
 FT  
 XX  
 PN  
 XX  
 PD  
 XX  
 XX  
 PF  
 XX  
 PR  
 XX  
 XX  
 PA  
 XX  
 XX  
 DR  
 DR  
 XX  
 PT  
 PS  
 XX  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 CC  
 SQ  
 Alignment Scores:  
 Pred. No.: 6,81e-36 Length: 1665  
 Score: 440.00 Matches: 134  
 Percent Similarity: 44.44% Conservative: 86  
 Best Local Similarity: 27.07% Mismatches: 171  
 Query Match: 20.51% Indels: 104  
 DB: 21 Gaps: 18  
 US-09-469-200d-2 (1-417) x AAZ88201 (1-1665)  
 QY 4 LeuLysAsnLeuIleThrValValAlaPheSerIlePhe-----Trp 17  
 Db 13 CTGACTACAGCCCTGCTGGTGGGCGACCATCTGTTGGCCCTGGTAGTGTGGGAGGC 72  
 QY 18 ValLeuLeuIleTyrValAsnValTyrLeuPhe-----GlyAlaLysGlySerLeuSer 35  
 Db 73 ATCCTGGCGGCTATGTGACAGGCTACCAAGTTATCCACAGAAAGGCACTACCTGTC 132  
 QY 36 -----IleTyrGlyPheLeuLeuIleAlaTyrLeuLeuValLysMetSerLeuSerPhe 53  
 Db 133 TTTGGCCCTCTACGGTGCATCTCTGGTGTCTACATCTCTCATCCAGAGCTGTTGGCTTC 192  
 QY 54 PheTyrLysProPheLysGlyArgAlaGlyGln----- 64  
 Db 193 CTGGAGCCCGCTCAATGCGGAGGCGGCGCCCTCAAGCTGCCTCTCCAGAGG 252  
 QY 65 ---TyrLysValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSerLeuLeuGlu 83  
 Db 253 TCSCGTTCAGTGGCACTCTGCTTGTCTTACCAAGAGAGCCCGAATACCTGCGCAG 312  
 QY 84 ThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrValValAspAsp 103  
 Db 313 TGCCTTCCTCAGTCAGCGCATTCCTTCCAAACCTCAAGTGTGTGCTAGTGGAT 372  
 QY 104 GlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAspThr----- 121  
 Db 373 GGCATTCGCCAGAA-----GATACCTACATGTGGAGATCTTCAT 414  
 QY 122 -----GlyAspLeuSerSerAsnValIleValHisArgSer-----Glu 134

/product= "HAS3"  
 /note= "modified hyaluronate synthase protein"

JP2000004886-A.

11-JAN-2000.

24-JUN-1998; 98JP-0193788.

24-JUN-1998; 98JP-0193788.

(SEKK) SEIKAGAKU KOGYO CO LTD.

WPI; 2000-140125/13.

P-PSDB; AAY68493.

A hyaluronate synthase modified protein - useful as a research reagent for biochemical research and medical development

Claim 18; Page 20-22; 30pp; Japanese.

The present invention describes a recombinant protein which consists of three continuous regions (N-terminal region, internal region and C-terminal region) where one or two regions among the above three regions is selected from the three hyaluronate synthase (HAS) modified proteins HAS1, HAS2 and HAS3 (e.g. a protein made from HAS1 amino acids 1 to 71, HAS2 amino acids 52 to 379 and HAS1 amino acids 410 to 583). The protein is useful as a research reagent for biochemical research and medical development. The invention provides a HAS modified protein of increased or lowered activity. The present sequence encodes mouse HAS3.

Sequence 1665 BP; 340 A; 453 C; 451 G; 421 T; 0 other;

ID	ABZ76736	standard; cDNA; 1665 BP.
XX		
AC	ABZ76736;	
XX		
DT	01-APR-2000	(first entry)
XX		
DE	Mouse hyaluronan synthase 3	encoding cDNA SEQ ID NO:3.
XX		
KW	Mouse; hyaluronan synthase;	HAS; HAS3; enzyme; ophthalmological;
KW	osteopathic; antiarthritic;	gene therapy; angiogenesis inhibitor;
KW	hyaluronic acid; dihydrazide;	bioconjugate; dry eye syndrome; eye;
KW	osteoarthritis; gene; ss.	
XX		
OS	Mus sp.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1665
FT		/tag= a
FT		/product= "hyaluronan synthase 3 (HAS3)"
XX		
FN	W02003006068-AL.	
XX		
PD	23-JAN-2003.	
XX		
PF	10-JUL-2001;	2001WO-US21785.
XX		
PR	10-JUL-2001;	2001WO-US21785.
XX		
PA	(CLEAR-) CLEAR SOLUTIONS BIOTECH INC.	
PI	Dehazya P, Chen W;	
XX		
WPI	2003-221664/21.	
DR	P-PSDB; ABP96030.	
XX		
PT	Novel dihydrazide derivatized	hyaluronic acid/nucleic acid bioconjugate
PT	for treating dry eye syndrome,	has derivatized hyaluronic acid
PT	crosslinked to nucleic acid	encoding protein with hyaluronan synthase
XX	activity -	
XX		
PS	Claim 19; Page 57-58;	62pp; English.
XX		
CC	The present invention describes a dihydrazide derivatised hyaluronic acid (HA)/nucleic acid (NA) bioconjugate (I), with derivatised HA linked to NA encoding protein with hyaluronan synthase (HAS) activity, where NA has sequence with 79-85% sequence identity (SI) to a sequence of 1752, 1659 or 1665 nucleotides (see ABZ76734 to ABZ76736) or a sequence encoding proteins with 66-90% SI to a sequence of 583, 552 or 554 amino acids (see ABP96028 to ABP96030). (I) has ophthalmological, osteopathic and antiarthritic activities, and can be used in gene therapy and as an inhibitor of angiogenesis, and as an inducer of expression of (HA) in human corneal epithelial cell. (I) is useful for transfecting a cell of an eye with nucleic acid, where the nucleic acid comprises (ABZ76734 to ABZ76736), by contacting the cell with (I) comprising the nucleic acid. (I) is useful for treating dry eye syndrome in an individual. (I) is useful in gene therapy applications for the treatment of a variety of medical conditions including dry eye syndrome or other medical conditions where an increase in the production of (HA) in the eye would be therapeutic (e.g., osteoarthritis of the articular joints). (I) is also useful for inhibiting angiogenesis for the treatment of macular degeneration or genes related to lipid biosynthesis that helps to restore the lipid component of the tear film, and as reagents for in vitro transformation of any cell, preferably a eukaryotic cell, more preferably a human eye cell. The present sequence encodes mouse HAS3 which is used in the exemplification of the present invention.	

Alignment Scores:	
Pred. No.:	6.81e-36
Score:	440.00
Percent Similarity:	44.44%
Best Local Similarity:	27.07%
Length:	1665
Matches:	134
Conservative:	86
Mismatches:	171

Query Match: 20.51% Indels: 104  
DB: 25 Gaps: 18  
US-09-469-2000d-2 (1-417) x ABZ76736 (1-1665)

QY 4 LeuLysAsnLeuIleThrValValAlaPheSerIlePhe-----Trp 17  
DB 13 CTGACACAGCCCTCGCTGGTGGGCGACCACTGCTGTTGCCCTGGTAGTGGCGAGGC 72  
QY 18 ValLeuLeuIleThrValValAlaPheSerIlePhe-----GlyAlaLeuGlySerLeuSer 35  
DB 73 ATCTGGGGGCTAAGTACAGAGCTACAGGCTTATCCACACAGAAAAGCACTACCTGTGC 132  
QY 36 -----IleTyrGlyPheLeuLeuIleAlaTyrLeuValLysMetSerLeuSerPhe 53  
DB 133 TTGGGCTCTACGGTGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 192  
QY 54 PheTyrLysProPheLysGlyArgAlaGlyGln-----64  
DB 193 CTGGAGCACCTCGAATGCGCAGGCGCGCCCTCAAGCTGCATCTGCCAGAGG 252  
QY 65 ---TyrLysValAlaIleThrProSerTyrAsnGluAspAlaGluSerLeuLeuGlu 83  
DB 253 TCGCTTCAGTGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 312  
QY 84 ThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrValValAspAsp 103  
DB 313 TGCCTTCGCTCAGTCCAGGCGATTCCTTCCAAACCTCAAGGTGGTCAATGCTAGTGGAT 372  
QY 104 GlySerAlaAspGluThrGlyLysArgIleGluAspTyrValArgAspThr-----121  
DB 373 GGCATCCCGAGNA-----GATACCTACATGTTGGACATCTTCCAT 414  
QY 122 -----GlyAspLeuSerSerAsnValIleValHisArgSer-----Glu 134  
DB 415 GAGGTGCTGGTGGCACTGAGCAAGCTGGCTTCTTGTGGCTGACATTTCCATGAG 474  
QY 135 LysAsnGlnGlyArgHisAla-----142  
DB 475 GCGGTGAAGGAG 534  
QY 143 -----GlnAlaTyrAlaPheGluArg-----149  
DB 535 GTGTGGGCGACACCTCTCATGCAATCATGCAAGTGGGGGCGCAAGGCTGAGTCAATG 594  
QY 150 -----SerAspAlaAspValPheLeuThrValAspSerAsp 161  
DB 595 TACACTGCTTCAAGGCCCTTGGCACTGAGTACATCCAGGTGTGCTGCTGCTGCTGCTGCTGCT 654  
QY 162 ThrTyrIleTyrProAspAlaLeuGluLeuLysThrPhe---AsnAspProThr 180  
DB 655 ACTGTGTGGAGCCCGCTGACCATTCCTCAACAGTATGATGATGATGATGATGATGATGATGATGAT 714  
QY 181 ValPheAlaAlaThrGlyHisLeuValArgAsnArgGlnThrAsnLeuLeuThrArg 200  
DB 715 GTAGAGGTGTGGAGGAGATGCTCAATCTCTCAACAGTATGATGATGATGATGATGATGATGATGAT 774  
QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThr 220  
DB 775 CTGACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 834  
QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValValPro 240  
DB 835 GCTGTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 894  
QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260  
DB 895 TTCTTGGAGGATGTTACATCAGAGTCTTCCAGCAGCAGAGTGGAGTGGAGTGGAGTGGAGTGGAG 954  
QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGly---LysThrValTyrGlnSerAla 279  
DB 955 CGGCACCTTACCAACCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1014

QY 280 LysCysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnAsnArgTyr 299  
DB 1015 AAGTGCCTCAG 1074  
QY 300 AsnLysSerPhePheArgLysIle-----308  
DB 1075 AGCAAGCTTACTTTCGGGAATGGCTCTACAATTCCTGTGTGGTTCATTAAGCAGCAGCTC 1134  
QY 309 ---IleSerValLysIleMetAsnAsn-----ProPheValAlaLeuThrIle 325  
DB 1135 TGGATGACCTATGAATCAGTGGTCCAGAGTTCCTTCCCATCTTCTCATGCTACAGT 1194  
QY 326 LeuGluValSerMetPheMetMetLeuValTyrSerValValAspPhePheValGlyAsn 345  
DB 1195 ATACAACCTT-----TTCTACCGTGGCGG 1218  
QY 346 ValArgGluPheAspTyrLeuArgValLeuAlaPheLeuValIleIlePheIleValAla 365  
DB 1219 ATC-----TGG---AACATTCCTCTCTCTGCTACAGAGATGATCTTCAATGTC 1263  
QY 366 LeuCysArgAsnIleHis---TyrMetLeuLysHisProLeuSerPheLeuLeuSerPro 384  
DB 1264 ATTATCAAGGCTACCTATGCTGCTCTCTGAGGCAATCGAGAGATGATCTTCAATGTC 1323  
QY 385 PheTyrGlyValIleHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThr 404  
DB 1324 CTCTACTCTCTCTATATGTCAGCCCTTTCGCCCAAGATCTTGTCTATTGCTATC 1383  
QY 405 IleArgAsnAlaAspTyrGlyThr-----ArgLysLysLeuLeu 417  
DB 1384 ATCAACAAGCTCTGGTGGGCACTTCTGCGAGAAACCATTTGTC 1428

RESULT 13  
ID ABLA1013 standard; cDNA; 1653 BP.  
AC ABLA1013;  
XX AC  
XX AC  
DT 30-JUL-2002 (first entry)  
XX  
XX Murine hyaluronan synthetase 2 (HAS2) cDNA fragment.  
XX Murine hyaluronan synthetase 2, HAS2; tumour; gene therapy; mouse; gene; ss.  
XX Mus sp.  
XX JP2002065280-A.  
XX  
XX 05-MAR-2002.  
XX  
XX 01-SEP-2000; 2000JP-0266260.  
XX  
XX 01-SEP-2000; 2000JP-0266260.  
XX  
XX (SEK) SEIKAGAKU KOGYO CO LTD.  
XX  
XX WPI; 2002-397661/43.  
XX  
XX A DNA vector for inhibiting the formation of a malignant tumor -  
XX  
XX Claim 1; Page 6; 9pp; Japanese.  
XX  
XX The invention relates to a DNA vector containing a 20 nucleotide sequence  
XX of the 3'-end of a mouse hyaluronan synthetase 2 (HAS2) cDNA sequence,  
XX or DNA hybridizable with a polynucleotide complementary to it under  
XX stringent conditions, and which can inhibit the translation of a  
XX hyaluronan synthetase 2 gene in which the transcription initiating point  
XX is recombined 5' upstream of the DNA. The vector is used for inhibiting  
XX the formation of a malignant tumour. The present sequence represents a  
XX mouse HAS2 cDNA fragment.  
SQ Sequence 1653 BP; 471 A; 387 C; 353 G; 442 T; 0 other;

PT never aminoguanidine derivatized nucleic acid/nucleic acid bioconjugate  
PT for treating dry eye syndrome, has derivatized hyaluronic acid crosslinked to nucleic acid encoding protein with hyaluronan synthase

XX

[illegible]

UD  
/42 ATTAAATTG-AACAAAACGGTAGCACTCTGCATCGCTGCGTACCAAGAGGACCCCTGACTA 800

Db 1857 CAGATGTTTGCATTTGCAACCATTAACAAAGCTGGTGGGGCACATCTGGAAGGAGAC 1916  
Qy 415 sLeuLeu 417  
Db 1917 CATTGTT 1923  
: : : : :  
: : : : :

Search completed: October 2, 2003, 12:56:38  
Job time : 698 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 2, 2003, 12:44:58 : Search time 82 Seconds

(without alignments)  
2244.595 Million cell updates/sec

Title: US-09-469-200D-2

Perfect score: 2145

Sequence: 1 MRTLKLTIVAPISIFVLL.....KLYSLTIRNADWCTRKLL 417

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09469200/runat\_02102003\_090348\_22797/app\_query.fasta.1.583  
-DB=Issued\_Patents\_NA -QW=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09469200.cgn1.1.56 @runat\_02102003\_090348\_22797 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA:\*  
1: /cgn2.6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2.6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2.6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2.6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2.6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2.6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1586.5	74.0	1512	2 US-08-270-581-1
2	1586.5	74.0	1512	4 US-09-146-893-1
3	463	20.6	536165	4 US-09-214-808-1
4	439	21.5	2947	4 US-08-675-499A-1
5	368.5	17.2	2117	3 US-09-155-768-1
6	366.5	17.1	2116	4 US-08-635-552A-1
7	242.5	11.3	4951	2 US-08-867-030B-5
8	242.5	11.3	4951	5 PCT-US95-06119-5
9	239.5	11.2	1326	4 US-09-107-532A-2775
10	239	11.1	1260	4 US-09-328-352-2963
11	234.5	10.9	1257	4 US-09-134-001C-1214
12	229	10.7	1263	4 US-09-328-352-1781

c	13	175	8.2	19390	4	US-08-961-527-86	Sequence 86, Appl
	14	166	7.7	4066	4	US-09-499-203-16	Sequence 16, Appl
	15	166	7.7	9321	4	US-09-499-203-1	Sequence 1, Appli
c	16	163	7.6	2673	4	US-09-252-991A-2993	Sequence 2993, Ap
	17	163	7.6	2715	4	US-09-252-991A-2686	Sequence 2686, Ap
	18	163	7.5	2814	4	US-09-252-991A-2875	Sequence 2875, Ap
	19	149	6.9	3084	1	US-08-551-437-1	Sequence 1, Appli
	20	149	6.9	3084	3	US-09-004-225-1	Sequence 1, Appli
	21	149	6.9	3084	3	US-09-084-346-1	Sequence 1, Appli
	22	149	6.9	3084	3	US-09-104-704-1	Sequence 1, Appli
	23	146.5	6.8	910	3	US-08-961-083-201	Sequence 201, App
	24	146.5	6.8	910	4	US-09-536-784-201	Sequence 201, App
	25	146	6.8	1311	4	US-09-328-352-290	Sequence 290, App
	26	144.5	6.7	1654976	4	US-08-916-421B-1	Sequence 1, Appli
	27	144	6.7	1830121	4	US-09-557-884-1	Sequence 1, Appli
	28	144	6.7	1830121	4	US-09-643-990A-1	Sequence 1, Appli
	29	143.5	6.7	912	4	US-09-252-991A-12584	Sequence 12584, A
c	30	140	6.5	8876	4	US-08-961-527-170	Sequence 170, App
	31	139.5	6.5	2166	4	US-09-107-532A-3235	Sequence 3235, Ap
	32	138	6.4	235	4	US-08-675-499A-25	Sequence 25, Appl
	33	137.5	6.4	522	4	US-08-858-207A-232	Sequence 232, App
c	34	137.5	6.4	1128	4	US-09-252-991A-1077	Sequence 1077, Ap
	35	137.5	6.4	1191	4	US-09-252-991A-1025	Sequence 1025, Ap
	36	137.5	6.4	3114	4	US-09-252-991A-977	Sequence 977, App
	37	135	6.3	235	4	US-08-675-499A-26	Sequence 26, Appl
	38	131	6.1	235	4	US-08-675-499A-23	Sequence 23, Appl
	39	128	6.0	506	4	US-09-252-991A-16525	Sequence 16525, A
	40	128	6.0	3060	4	US-09-115-150-1	Sequence 1, Appli
	41	127	5.9	951	4	US-09-107-532A-1037	Sequence 1037, Ap
	42	127	5.9	14602	1	US-08-597-236-1	Sequence 1, Appli
	43	127	5.9	14602	1	US-08-746-682A-1	Sequence 1, Appli
c	44	124.5	5.8	14187	4	US-09-453-702B-121	Sequence 121, App
	45	124	5.8	1026	4	US-09-328-352-3052	Sequence 3052, Ap

## ALIGNMENTS

### RESULT 1

US-08-270-581-1  
; Sequence 1, Application US/08270581  
; Patent No. 5856168  
; GENERAL INFORMATION:  
; APPLICANT: Weigel, Paul H.  
; APPLICANT: Deangelis, Paul L.  
; APPLICANT: Papaconstantinou, John  
; TITLE OF INVENTION: Hyaluronate Synthase Gene and Uses  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P. O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/270,581  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: UTSG:161\PAR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713-787-1400  
; TELEFAX: 713-789-2679  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1512 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 193..1449  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1488..1510  
 ; US-08-270-581-1

## Alignment Scores:

Pred. No.: 4,66e-187 Length: 1512  
 Score: 1586.50 Matches: 298  
 Percent Similarity: 85.19% Conservatives: 53  
 Best Local Similarity: 72.33% Mismatches: 60  
 Query Match: 73.96% Indels: 1  
 DB: 2 Gaps: 1

US-09-469-200D-2 (1-417) x US-08-270-581-1 (1-1512)

QY 5 LysAsnLeuLeuThrValValAlaPheSerIlePheTrpValLeuLeuLeuTyrValAsn 24  
 DB 205 AAAAAAAGCTTAATGTTTATTCCTTTATTTTGTATATCTATCTGTGATTAAT 264  
 QY 25 ValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeuLeuLeuLeuTyr 44  
 DB 265 AGTATCTATTGGACAC---TCAACTGTAGGAGTATGAGGTAATATTATAACCTAT 321  
 QY 45 LeuLeuValLysMetSerLeuSerPhePheTyrLysProPheLysGlyArgAlaGlyGln 64  
 DB 322 CTAGTTATCAAACTTGGATTCTTCCTTTATGAGCCATTTAAAGGAATCCCATGAC 381  
 QY 65 TyrLysValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSerLeuLeuGluThr 84  
 DB 382 TATAAAGTGTCTGCTTAATCTCTTATTAATGAAGATGCCGAGTCATTATTAGAACA 441  
 QY 85 LeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrValValAspAspGly 104  
 DB 442 CTTAAAGTGTGTTACACAGACCTATCCGTTATCAGAAATTTATATGTCATGATGGG 501  
 QY 105 SerAlaAspGluThrGlyLysArgIleGluAspTyrValArgAspThrGlyAspLeu 124  
 DB 502 AGTTCAACACAGATGCAATCAATTAATGAAGATGTAATAAGAGATGATAT 561  
 QY 125 SerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArgHisAlaGlnAla 144  
 DB 562 TGTCAAAAGCTTATCTTCACGTTCCCTGTCATTAAGGAAGAACCCATGCTCAAGCG 621  
 QY 145 TrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSerAspThrTyrIle 164  
 DB 622 TGGCATTGAAAGATCTCAGCGTACGCTTTTAAACCGTAGACTCAGATCTATTATC 681  
 QY 165 TyrProAspAlaLeuGluLeuLeuLysThrPheAsnAspProThrValPheAlaAla 184  
 DB 682 TATCAACGCTTATGAGAGACTCTCTAAAGAGCTTCAATGATGAGACCTTTATGCTGCA 741  
 QY 185 ThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIle 204  
 DB 742 ACAGGACATTGATCTAGAAACAGACAACTATCTATTACCGGACTTACAGATC 801  
 QY 205 ArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThrGlyAsnIleLeu 224  
 DB 802 CGTTACGATAATGCCCTTGGGTGGAGCGTCTGCTCAATCAATTAACAGGTAATTTTA 861  
 QY 225 ValCysSerGlyProLeuSerValTyrArgArgGluValValProAsnIleAspArg 244  
 DB 862 GTTGTCTCAGGACCATGAGTATTATCGACGTGAAGTGATTTATCTCAACTAGAGCG 921  
 QY 245 TyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThr 264

Db 922 TATATAAATCAAAATTCCTAGGTTTACCTGTTAGCATTTGGGGATGATCGATGTTTAAACA 981  
 QY 265 AsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLysCysIleThrAsp 284  
 Db 982 AATTATGCTATTGATTAGGACGACTGCTACCAATCAACAGCTAGATGATGATGAT 1041  
 QY 285 ValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTyrPAsnLysSerPhePhe 304  
 Db 1042 GTACCTTCCCAATTAAGAGTTATTTAAAGCAACAATCGATGGAATAAATCTTTTTT 1101  
 QY 305 ArgGluSerIleLeuSerValLysLysIleMetAsnAsnProPheValAlaLeuTyrThr 324  
 Db 1102 AGAGATCTATTATTTCTTAAATAAATTTCTTAATCCCATCGTTCCTTATGGACT 1161  
 QY 325 IleLeuGluValSerMetPheMetMetLeuValTyrSerValValAspPhePheValGly 344  
 Db 1162 ATTTTCGAAGTCGTTATGTTTATGATGTTGATGTTGTCGAATGGCAATCTTTGTTTAA 1221  
 QY 345 AsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIleIlePheIleVal 364  
 Db 1222 CAAGCTATTCATTAAGCTTATTAACCTTTTTCCTTTTATCCATCATCTTTATCGGT 1281  
 QY 365 AlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPheLeuLeuSerPro 384  
 Db 1282 GCITTTATGCTGTAATGTTTATATATGTTCAACATCTGCTAGTTTGTGTTATCTCT 1341  
 QY 385 PheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThr 404  
 Db 1342 CTGTATGGAATATACACTTGTGTTCTTACAGCCCTAAACCTTTATCTTTATGACCC 1401  
 QY 405 IleArgAsnAlaAspTrpGlyThrArgLysLysLeu 416  
 Db 1402 ATTAAATACGGAATGGGACACGTAAGAAGTC 1437

## RESULT 2

US-09-146-993-1  
 ; Sequence 1, Application US/09146893  
 ; Patent No. 6455304  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weigel, Paul H.  
 ; DeAngelis, Paul L.  
 ; Papaconstantinou, John  
 ; TITLE OF INVENTION: Ryaluronate Synthase Gene and Uses  
 ; Thereof  
 ; NUMBER OF SEQUENCES: 7  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Arnold, White & Durkee  
 ; STREET: P. O. Box 4433  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: USA  
 ; ZIP: 77210-4433  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/146,893  
 ; FILING DATE: 03-Sep-1998  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/270,581  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Parker, David L.  
 ; REGISTRATION NUMBER: 32,165  
 ; REFERENCE/DOCKET NUMBER: UTSG:161\PAR  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 713-787-1400  
 ; TELEFAX: 713-789-2679  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:

;  
 ; LENGTH: 1512 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 193..1449  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1488..1510  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 US-09-146-893-1

Alignment Scores:  
 Pred. No.: 4,66e-187 Length: 1512  
 Score: 1586.50 Matches: 298  
 Percent Similarity: 85.19% Conservative: 53  
 Best Local Similarity: 72.33% Mismatches: 60  
 Query Match: 73.96% Indels: 1  
 DB: 4 Gaps: 1

US-09-469-200D-2 (1-417) x US-09-146-893-1 (1-1512)

QY 5 LysAsnLeuIleThrValValAlaPheSerIlePheTrpValLeuIleTyrValAsn 24  
 DB 205 AAAAAAATTAAATGTTTATCTTTATTTTATGATATCTATCTGATTATCTAAAT 264  
 QY 25 ValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeuIleAlaTyr 44  
 DB 265 ATGTATCTATTGGAGACA---TCAACTGTAGGAATTTATGGAGTAATATTAATACCTAT 321  
 QY 45 LeuLeuValLysMetSerLeuSerPhePheTyrLysProPheLysGlyArgAlaGlyGln 64  
 DB 322 CTAGTTATCAACTGGATTATCTTCTTTATGAGCAATTAAGGAAATCCACATGAC 381  
 QY 65 TyrLysValAlaAlaIlelleProSerTyrAsnGluAspAlaGluSerLeuLeuGluThr 84  
 DB 382 TATAAAGTGTCTGTGTAATCTCTTATAATGAAGTCCGAGTCATATTAGAAACA 441  
 QY 85 LeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrValValAspGly 104  
 DB 442 CTAAAGTGTGTAGCAGACCTATCGTATCAGAAATTTATTTGTTGATGG 501  
 QY 105 SerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAspThrGlyAspLeu 124  
 DB 502 AGTCAACACAGATGCAATACAAATTAATGAAGATGTAATAGAGAGTGGATAT 561  
 QY 125 SerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArgHisAlaGlnAla 144  
 DB 562 TGTCGAACGTTATCGTTCACCGTTCCTTCTCAATAAAGAAACGCCATGCTCAAGG 621  
 QY 145 TrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSerAspThrTyrIle 164  
 DB 622 TGGGCATTTGAAGATCTGACGTCACGTTTTTAAACGCTAGACTCAGATCTATATC 681  
 QY 165 TyrProAspAlaLeuGluGluLeuLeuLysThrPheAsnAspProThrValPheAlaAla 184  
 DB 682 TATCCAAATGCTTAGAAGAACTCTCTAAAGCTCAATGATGAGACAGTTTATGTCGA 741  
 QY 185 ThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIle 204  
 DB 742 ACAGGACATTTGAATGCTAGAACACAGACAACTAATCTATTACCGCACTTACAGATATC 801  
 QY 205 ArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThrGlyAsnIleLeu 224  
 DB 802 CGTTACGATTAATGCTTTGGGTGGAGCGTCTCTCAATCAATTAACAGGTATATTTA 861  
 QY 225 ValCysSerGlyProLeuSerValTyrArgGluValValValProAsnIleAspArg 244  
 DB 862 GTTTGCTCAGACCACTGAGTATTATCGAGTGAAGTATTATTCCTACTTACCTAGAGCC 921  
 QY 245 TyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspArgCysLeuThr 264

DB 922 TATAAAATCAACATCTCTAGGTTTACCTGTAGCATTTGGGATGATCGATGTTTAAACA 981  
 QY 265 AsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLysCysIleThrAsp 284  
 DB 982 AATATGCTATTGATTAGGACCACTGCTTACCAATCAACAGCTAGATGCTACTGAT 1041  
 QY 285 ValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsnLysSerPhePhe 304  
 DB 1042 GTACCTTTCCAATTAAGAGTTATTAAAGCAACAATAATCGATGAATAATCTTTTTT 1101  
 QY 305 ArgGluSerIleLeuSerValLysIleMetAsnAsnPropheValAlaLeuThrPhe 324  
 DB 1102 AGAGATCTATATTTCTTTAAATAATTTCTTTTAAATCCCATCGTGCCTTATGGACT 1161  
 QY 325 IleLeuGluValSerMetMetMetLeuValTyrSerValValAspPhePheValGly 344  
 DB 1162 ATTTTCGAAGTCGTTATGTTATGATGTTGATGTCGCAATTTGGAAATCTTTTGTAT 1221  
 QY 345 AsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIleIlePheIleVal 364  
 DB 1222 CAAGCTATTCAATTAGACCTTATTAAACTTTTGGCTTTTATCCATCATCTTTATCGTT 1281  
 QY 365 AlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPheLeuLeuSerPro 384  
 DB 1282 GCTTTATGCTGTAATGCTTATGCTTATGCTTATGCTCAACATCTCTAGTTTGTATCTCCT 1341  
 QY 385 PheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThr 404  
 DB 1342 CTGTATGGAATATTACACTGTTGCTTACAGCCCTTAAACTTTATCTTTATGCAACC 1401  
 QY 405 IleArgAsnAlaAspTrpGlyThrArgLysLysLeu 416  
 DB 1402 ATTAAATACGAATGGGACACGCTAAAGAGGTC 1437

## RESULT 3

US-09-214-808-1/C  
 ; Sequence 1, Application US/09214808A  
 ; Patent No. 6475793  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosenthal, Andre  
 ; APPLICANT: Freiberg, Christoph  
 ; APPLICANT: Perret, Xavier Philippe  
 ; APPLICANT: Broughton, William John  
 ; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic  
 ; Patent No. 6475793  
 ; TITLE OF INVENTION: Plasmid  
 ; FILE REFERENCE: CARP0068  
 ; CURRENT APPLICATION NUMBER: US/09/214,808A  
 ; CURRENT FILING DATE: 1999-06-22  
 ; PRIOR APPLICATION NUMBER: PCT/IB97/00950  
 ; PRIOR FILING DATE: 1997-07-10  
 ; NUMBER OF SEQ ID NOS: 1  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 536165  
 ; TYPE: DNA  
 ; ORGANISM: Rhizobium  
 US-09-214-808-1

Alignment Scores:  
 Pred. No.: 3,47e-43 Length: 536165  
 Score: 463.00 Matches: 131  
 Percent Similarity: 48.75% Conservative: 64  
 Best Local Similarity: 32.75% Mismatches: 185  
 Query Match: 21.59% Indels: 20  
 DB: 4 Gaps: 10

US-09-469-200D-2 (1-417) x US-09-214-808-1 (1-536165)

QY 27 LeuPheGlyAlaLysGly-----SerLeuSerIleTyrGlyPheLeuLeuIleAlaTyr 44  
 DB 157330 CTGCTTGGCAGCGCGCGGTAGCCATCTCCTTGTATGAGCACTCTCGACGCGCTTAC 157271

QY 45 -----LeuLeuValLysMetSerLeuSerPhePheTyrLysProPhe 58  
Db 157270 AAGGCGATGCGAGCTATATACGCTTGCACAAACACACAGCTGCGGTCAAGCGCGCGT 157211

QY 59 LysGlyArgAlaGlyGlnTyrLysValAlaAlaIleProSerTyrAsnGluAspAla 78  
Db 157210 ACCGGCTCGGTGCGCGGAGCGTGTATGCTACCTCGCTACATGAGATCGG 157151

QY 79 GluSerLeuGluThrLeuLysSerValGlnGlnGlnThrTyrPro----LeuAlaGlu 97  
Db 157150 CGGGCGCTCTCGCGCTGCTAGCTTCCATTGCAAGCAGGACTACGCTGAGAGTTCGGG 157091

QY 98 IleTyrValAlaAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyr 117  
Db 157090 GTCTAGTGGTGGACGAGCTTCTGCAATCGCAAGCCATCTACCTGTACAGATCAT 157031

QY 118 ValArgAspThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGln 137  
Db 157030 TAT-----GGTGGCGACCGAGGTTCGGCTTTATCTCTG-----ATGCCAAGACGTC 156983

QY 138 GlyLysArgHisAlaGlnAlaTrpAlaPheGluArgSerAspAlaAspValPheLeuThr 157  
Db 156982 GGAAGCGCAGGCGCGAGATTCCTCGGATACGGGAATCATCGGAGATTTGGTGTCAAC 156923

QY 158 ValAspSerAspThrTyrIleTyrProAspAlaLeuGluGluLeuLeuLysThrPheAsn 177  
Db 156922 GTTGACTCGGACAGCACCATTTCGCCGAGCTAGTCACGAACTTCGCCGTGAAGATGTAC 156863

QY 178 AspProThrValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeu 197  
Db 156862 AGTCCCGCGGTGGCGCGCGATGGTTCAGTTCAGCGCCAGCAACCGCAGCACATGG 156803

QY 198 LeuThrArgLeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGln 217  
Db 156802 CTGAGCGGTGTATCATGATGAGTACTGGCTCGCTCGCAGGAGGACCGCGACAG 156743

QY 218 SerValThrGlyAsnLeuValCysSerGlyProLeuSerValTyrArgGluVal 237  
Db 156742 GCTCGCTTTGGAGCGGTATGTGTGTCGGCGCGGTGTGCCATGTACCGCGGTCGCA 156683

QY 238 ValValProAsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIle 257  
Db 156682 CTCCATTGCTGCTGATTAATACGAGACCACTGTTTCGAGCGCCAGCGCACTTC 156623

QY 258 GlyAspAspArgCysLeuThrAsnTyrAlaThrAspLeuGly---LysThrValTyrGln 276  
Db 156622 GGGGAAGCGCCGACCTCACATCTCATCTGATGAGCGCTTTCGACCGAGTACGTT 156563

QY 277 SerThrAlaLysCysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGln 296  
Db 156562 CGGAAGCCATCGCGCGAGCGGTGTTCCAAACTCGATGGCGCTATCTGCGCCACAA 156503

QY 297 AsnArgTyrAsnLysSerPheArgGluSerIleIleSerValLysLysIleMetAsn 316  
Db 156502 CTGCGCTGGGCGAGCAGCGTTTCGGGACACATTCGCGCTCGGCTACTGCGCGGCG 156443

QY 317 ---AsnProPheValAlaLeuThrThrIleLeuGluValSerMetPheMetLeuVal 335  
Db 156442 CTGTATGCTATCTTACGCTGAGCTGATCGGACAGATCTTGGTCCGCTGCTCTAGCC 156383

QY 336 TyrSerValValAspPhePhe-----ValGlyAsnValArgGluPheAspTrpLeuArg 353  
Db 156382 CTCTCGGTCTGAGCGGGCTAGCACAGCTCGCTCTGAGCGCCACAGTTCGCTGTCGAGC 156323

QY 354 ValLeuAlaPheValIleIlePheIleValAlaLeuCysArgAsnIleHisTyrMet 373  
Db 156322 ATCTGT-----ATGATTGCATCTATGACAAATGGTCCGCTGC---GGCGTGGCGCGTT 156272

QY 374 LeuLysHisProLeuSerPheLeuLeuSerProPheTyrGlyValLeuHisLeuPheVal 393  
Db 156271 CGAGCGGAGAGCTGGATTCCTTGGGTTTTCGCTGCGACACCTCTCTCAACGCTGCTCTC 156212

QY 394 LeuGlnProLeuLysLeuTyrSerLeuPheThrIleArgAsnAlaAspTrpGlyThrArg 413  
Db 156211 CTGTCCCGCTCAAAAGCATATGCGTGTGTCAGCTTGAGCAACAGCGACTGCTGCGGT 156152

RESULT 4  
US-08-675-499A-1  
; Sequence 1, Application US/08675499A  
; Patent No. 6492150  
; GENERAL INFORMATION:  
; APPLICANT: McDonald, J. A.  
; APPLICANT: Spicer, A. P.  
; APPLICANT: Augustin, M. L.  
; TITLE OF INVENTION: GENE ENCODING HYALURONAN  
; TITLE OF INVENTION: SYNTHASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.  
; STREET: P.O. Box 2938  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/675,499A  
; FILING DATE: 03-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Woessner, Warren D  
; REGISTRATION NUMBER: 30,440  
; REFERENCE/DOCKET NUMBER: 150.1700S1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-359-3260  
; TELEFAX: 612-359-3263  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2947 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-675-499A-1

Alignment Scores:  
Pred. No.: 5,99e-44 Length: 2947  
Score: 439.00 Matches: 134  
Percent Similarity: 46.65% Conservative: 82  
Best Local Similarity: 28.94% Mismatches: 183  
Query Match: 20.47% Indels: 65  
DB: 4 Gaps: 14

US-09-469-200D-2 (1-417) x US-08-675-499A-1 (1-2947)

QY 7 LeuIleThrValValAlaPheSerIlePheTrpValLeuLeuIleTyrValAsnValTyr 26  
Db 574 CTCTCGGAATCACAGCTGCTATATTGTTGGCTACCATTTTTCACCAACAGATAATTAC 633

QY 27 LeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeuLeuIleAlaTyrLeuLeu 46  
Db 634 TACTTC-----TCATTGGAGCTGTACGGTGCCTTTTGGAGCTGCGCATCTCATC 681

QY 47 Val-LysMetSerLeuSerPhePhe-----TyrLysProPheLysGlyArg 61  
Db 682 ATCCAAAGCTCTTTGCTTTTGGAGCAACCGGAAATGAGAGATGCTCTTGAACCCCG 741

```

QY 61 gAlaGlyGlnTyrLys---ValAlaAlaIlelleProSerTyrAsnGluAspAlaGluSe 80
Db 742 ATTAATG-ACAAACGGTAGCACTCTGCATCGCTGCGTACCAGAGAGCCCTGACTA 800
QY 80 rLeuLeuGluThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVa 100
Db 801 CTACGGAATGTTTGAATCTGTGAAGAGCTGACCTACCTGGGATTAAAGTCGTGAT 860
QY 100 lValAspAspGlySerAlaAsp-----107
Db 861 GGTCATCGATGGGAACCTCAGACGACGACCTTTACATGATGGAATATTACCGAAGTTAT 920
QY 108 -----GluThrGlyLeuLys 112
Db 921 TGCAGGAGCAAAATCGCCACGTACATCTGGAAGAACAACTTTCATGAAAGAGGACCTGG 980
QY 112 sArgIleGluAspTyrValArgAspThrGlyAsp-----LeuSerSerAs 127
Db 981 TGAGACAGAAGAGTCCCATTAAGAAAGTTTCAACATGTCAACCAATTGGTCTGTCTAA 1040
QY 127 nValIleValHisArgSerGluLysAsnGlnGlySerHisAlaGlnAlaTrpAlaPh 147
Db 1041 CAAAAGTATTTGCATCATGCAAAATGGGGTGGAAAGAGAGAGTCAATGTACACAGCCTT 1100
QY 147 eGluArg-----SerAspAlaAspValPheLeuThrValAspSerAspThrTyrIle 165
Db 1101 CAGAGCACTGGGGCAAGCGTGCATTTATGATGAGTGTGTGACTCAGATACATGCTTGA 1160
QY 165 rProAspAlaLeuGluGluLeuLeuLysThrPhe---AsnAspProThrValPheAla 184
Db 1161 CCTGTGCTCATCTGTGAGATGTGAAGTCTTAGAGAGAGACCTATGTTGGTGGAGTGT 1220
QY 184 aThrGlyHisLeuAsnValArgAsnGlnThrAsnLeuLeuThrArgLeuThrAspI 204
Db 1221 TGGAGGAGATGCCAGATTTTAAACAAGTATGATCTGATCTCTTCTTCAGCAGAGCT 1280
QY 204 eArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThrGlyAsnIle 224
Db 1281 GAGATACTGGATGGCTTTAATATAGAAGGGCTGCCAGTCTTATTTGGCTGTGTCCA 1340
QY 224 uValCysSerGlyProLeuSerValTyrArgGluValValValProAsnIleAsp 244
Db 1341 GTCCATAACCGGCTCTCTGGGAATGTACAGAACTCCTTGTGTCATGAATTTGTGAAGA 1400
QY 244 gTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeu 264
Db 1401 CTGGTACATCAGGAATTCATGGTAACCAATGCAGTTTGGTGCAGCAGGACACTTAC 1460
QY 264 rAsnTyrAlaThrAspLeuGly---LysThrValTyrGlnSerThrAlaLysCysIle 283
Db 1461 CAACAGGTTGTGAGTCTGGGCTATGCAACTAAATACAGGCTCGGTCGAAGTGCCTTAC 1520
QY 283 rAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsnLysSer 303
Db 1521 TGAATCCCTCCATAGATATCTGAGATGCTGAACACGACGACCCGATGGAGCAGTCTA 1580
QY 303 ePheArgGluSerIleSerValLysIleMetAsnAsnProPheValAlaLeuTr 323
Db 1581 CTTCGAGAGTGGTGTACAAATGCCATGTGTTTCAACAGCAT-----CACCTGTG 1631
QY 323 pThrIleLeuGluValSerMet-----PheMetMetLeuValTyrSerVa 338
Db 1632 GATGACCTATGAAGCTGTATACATGATCTTTCTCTTCTTCATTCACACATCAT 1691
QY 338 lValAspPheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLe 358
Db 1692 CCAGCTCTTCTACAGGGGTAAATC-----TGG---AACATCTCTCTCTCTCT 1736
QY 358 uValIleIlePheIleValAlaLeuCysArg---AsnIleHisTyrMetLysHisPr 377
Db 1737 GTTAACGTGCCAGTACTGGGTCTCATCAAGTCACTTTTGGCCAGCTGCCCTAGAGAAA 1796
QY 377 oLeuSerPheLeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnPro 397

```

```

Db 1797 TATCTCATGTTATTCATCTCTCTGTTATTCAGTGTATATACATGCTACTTCTCTGC 1856
QY 397 ulysLeuTyrSerLeuPheThrIleAlaGlnAlaAspTrpGlyThr-----ArgLysLy 415
Db 1857 CAAAGATGTTTGCATTTGCAATTTGCAACCAATTAACAAGCTGGTGGGACATCTGGGAAGGAGAC 1916
QY 415 sleuLeu 417
Db 1917 CATTTGTT 1923

RESULT 5
US-09-155-768-1
; Sequence 1, Application US/09155768A
; Patent No. 6162908
; GENERAL INFORMATION:
; APPLICANT: SEIKAGAKU CORPORATION
; TITLE OF INVENTION: POLYPEPTIDE OF HYALURONAN SYNTHASE OF
; TITLE OF INVENTION: HUMAN-ORIGIN AND DNA ENCODING THE SAME
; FILE REFERENCE: TOYAM37.001APC
; CURRENT APPLICATION NUMBER: US/09/155,768A
; EARLIER FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: JP 8-084326
; EARLIER FILING DATE: 1996-04-05
; EARLIER APPLICATION NUMBER: JP 8-109663
; EARLIER FILING DATE: 1996-04-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2117
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (149)...(1777)
US-09-155-768-1

Alignment Scores:
Pred. No.: 2e-35 Length: 2117
Score: 368.50 Matches: 137
Percent Similarity: 42.08% Conserved: 73
Best Local Similarity: 27.45% Mismatches: 182
Query Match: 17.18% Indels: 107
DB: Gaps: 20

US-09-469-200D-2 (1-417) x US-09-155-768-1 (1-2117)

QY 4 LeuLysAsnLeuIleThrValValAlaPheSerIlePheTrpValLeuLeuIleTyrVal 23
Db 98 CTGGCCCGGAGGGTGTGCACCATCGCTTCGCCCTGCTC---ATCTGCCCTCATGACC 154
QY 24 AsnValTyrLeuPheGlyAlaLys-----GlySerLeuSer----- 35
Db 155 TGGGCGCTACCGCCCGCGGGTCCGCTGCGCTCGATCGCTACGCGCTCTGGCTTCGGC 214
QY 36 lIleTyrGlyPheLeuLeuIleAlaTyrLeuLeuValLysMetSerLeuSerPhePheTyr 55
Db 215 CTCTACCGGGCTCTCTTTCAGCCAC---CTGGTGGGCGAGAGCCTC---TTCGCTAC 268
QY 56 LysProPheLysGlyArgAlaGlyGlnTyrLys----- 66
Db 269 CTGGAGCACCGCGGGTGGCGGGCGCGGGCGCGCGTGGATGATGACACCGCCGCGC 328
QY 67 ---ValAlaAlaIlelleProSerTyrAsnGluAspAlaGluSerLeuLeuGluThrLeu 85
Db 329 AGTGTGGCGTGCACCATCTCGGCTACAGAGAGACCCCGCTACCTGGCCAGTGCCTG 388
QY 86 LysSerValGlnGlnThrTyrProLeuAla----- 96
Db 389 CGGTCCCGCGCCCTCTCTACCGCGCGCGCGCGCTGCGCTGCTCATGTTGGTGGT 448
QY 97 -----GluIleTyrValValAsp----- 102

```

Db 449 GGCAACCGCGGAGGACCTCTACATGTGACATGTTCCGGGAGGCTCTTCGTCGACGAG 508  
 QY 103 -----AspGlySer----- 105  
 Db 509 GACCGCGCACATGCTGTGGGACGCACTACTACACAGCCCTGGGAACCGCGGCGGCG 568  
 QY 106 ---AlaAspGluThrGlyLeuArgAspGlyLeuArgAspGlyLeuArgAspGlyLeu 124  
 Db 569 GCGCGCTGGGCGGCGGAGCTATCG--GAGGTGAGCGGAGATCTCGGCGGCTG 625  
 QY 125 SerSerAsnValIleValHisArg-----SerGluLysAsnGlnGly 138  
 Db 626 GCGTGGAGGCGGCTGTGGAGGACTCCAGAGTGGCTGTGGCGGCGAGCGTGGGCGGCG 685  
 QY 139 LysArgHisAlaGlnAlaTrpAlaPheGluArg-----SerAspAlaAspValPheLeu 156  
 Db 686 AGCGCGAGTCATGTACACAGCTTCAGCGCTCGGAGATTCGGTGGAGTACGTCGACG 745  
 QY 157 ThrValAspSerAspThrTrpIleThrProAspAlaLeuGluLeuLeuLysThrPhe 176  
 Db 746 GTCTGTGCTGCTCGACACAGGTTGGACCCCATGCGTGTGGAGCTCGTGGGCTGCTG 805  
 QY 177 Asn---AspProThrValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThr 195  
 Db 806 GACGAGGACCCCGGAGGCTGTGGTGGGATGTCGGATCTTAACCCCTCTGGAC 865  
 QY 196 AsnLeuLeuThrArgLeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAla 215  
 Db 866 TCCTGGTCTGAGTCTTCAACAGCGCTGGATCTGGTGGTCTTCAATGTGGAGCGGT 925  
 QY 216 AlaGlnSerValThrGlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgArg 235  
 Db 926 TGTGACGAGTACTTCCACTGTATGCTATGCTACAGCGGCTCTCTAGGCTATATAGGA 985  
 QY 236 GluValValProAsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProVal 255  
 Db 986 AACCTCTCCAGCATCTTCTGAGCGCTGTACAGCAGGATCTCTGGTGGTCTGCTG 1045  
 QY 256 SerIleGlyAspArgCysLeuThrAsnTyrAlaThrAspLeuGly---LysThrVal 274  
 Db 1046 ACTTTGGGATGACGGCGACCTTCAACACCGCATGCTCAGCATGGTATGCTACCAAG 1105  
 QY 275 TyrGlnSerThrAlaLysCysIleThrAspValProAspLysMetSerThrTyrLeuLys 294  
 Db 1106 TACACTCCAGTCCCGCTGCTACTACAGACGCGCTCTCTCTCTCTCTCTCTCTCTCT 1165  
 QY 295 GlnGlnAsnArgTyrAsnLysSerPheArgGluSerIle----- 308  
 Db 1166 CAGCAGACAGCTGTCCAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1225  
 QY 309 -----IleSerValLysIleMetAsnAsn-----ProPheVal 320  
 Db 1226 CACCGGCAACATGCTGGATGAGTACAGAGCGGTGCTCTCGGCGCTGCTCTCTCTCT 1285  
 QY 321 AlaLeuTriPheLeuGluValSerMetPheMetMetMetMetMetMetMetMetMet 340  
 Db 1286 GTGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309  
 QY 341 PhePheValIleGlyAsnValArgGluPheAspThrPheArgValLeuAlaPheLeuValIle 360  
 Db 1310 TTCTACGCGGCGGCGCTTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1354  
 QY 361 IlePheIleValAlaLeuCysArg---AsnIleHisTyrMetLeuLysHisProLeuSer 379  
 Db 1355 GTGACGCGGCTGGCACTGGCAAGCGGCTCTCGCGGCTGCTGCTGCTGCTGCTGCTGCT 1414  
 QY 380 PheLeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeu 399  
 Db 1415 ATGTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1474  
 QY 400 TyrSerLeuPheThrIleArgAsnAlaAspThrPheThr-----ArgLysLysLeu 416  
 Db 1475 CTGGCGCTAGTACCACTGACACAGAGTGGCTGGGCGACCTCTCGGCGCGCGGAGCTG 1531

RESULT 6  
 US-08-635-552A-1  
 ; Sequence 1, Application US/08635552A  
 ; Patent No. 6423514  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Briskin, Michael J.  
 ; TITLE OF INVENTION: Mammalian Hyaluronan Synthases, Nucleic  
 ; TITLE OF INVENTION: Acids and Uses Thereof  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
 ; STREET: Two Militia Drive  
 ; CITY: Lexington  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02173  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/635,552A  
 ; FILING DATE: 22-APR-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brook, David E.  
 ; REGISTRATION NUMBER: 22,592  
 ; REFERENCE/DOCKET NUMBER: LKS95-07  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-861-6240  
 ; TELEFAX: 617-861-9540  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2116 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 36..1769  
 ; US-08-635-552A-1  
 Alignment Scores:  
 Pred. No.: 3,55e-35 Length: 2116  
 Score: 366.50 Matches: 135  
 Percent Similarity: 42.12% Conservative: 76  
 Best Local Similarity: 26.95% Mismatches: 179  
 Query Match: 17.09% Indels: 111  
 DB: 4 Gaps: 19  
 US-09-469-200D-2 (1-417) x US-08-635-552A-1 (1-2116)  
 QY 4 LeuLysAsnLeuIleThrValValAlaPheSerIlePheTrpValLeuLeuIleTyrVal 23  
 Db 90 CTGGCCCGGAGGCTGCTGACCATCGCTTCGCCGCTG-----CTCATCTGGGCTC 140  
 QY 24 AsnValTyrLeuPheClyAlaLysGlySerLeuSer----- 35  
 Db 141 ATGACCTGGCGCTACGCGCGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 200  
 QY 36 -----IleTyrClyPheLeuLeuIleAlaTyrLeuLeuValLysMetSerLeuSerPhe 53  
 Db 201 TTCGGCTCTGAGGGCTTCTCTTTCAGGCGAC---CTGGTGGCGAGAGCTC---TTC 254  
 QY 54 PheTyrLysProPheLysGlyArgAlaGlyGlnTyrLys----- 66  
 Db 255 GCGTACCTGGAGCAGCGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 314  
 QY 67 -----ValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSerLeuLeuGlu 83  
 Db 1475 CTGGCGCTAGTACCACTGACACAGAGTGGCTGGGCGACCTCTCGGCGCGCGGAGCTG 1531

Db 315 CGCGCAGTGTGGCGCTGACCATCTCCGCTACACGAGGAGACCCCGGTACTCGCGCCAG 374  
Qy ThrLeuLysSerValGlnGlnGlnThrTyrProLeuAla----- 96  
Db 375 TGCCTGGCGTCCGCGCGCGCTGCTGTACCGCGCGCGCTGCGGTGCTCATGGTG 434  
Qy 97 -----GluIleTyrValValAsp----- 102  
Db 435 GTGATGGCAACCGCGCGGAGGACCTTACATGTGTCACATGTTCCGCGAGGTCTTGCT 494  
Qy 103 -----AspGlySer----- 105  
Db 495 GACGAGACCCCGCCACCTAGCTGTGGGACGCACTACACACGCGCTGGGAACCGCG 554  
Qy 106 -----AlaAspGluThrGlyLysArgIleGluAspTyrValArgAspThrGly 122  
Db 555 CGCGCGGCGGTGGCGCGGAGGCTATCGG-----GAGGTGGAGCGGAGGATCTGGG 611  
Qy 123 AspLeuSerSerAsnValIleValHisArg-----SerGluLysAsn 136  
Db 612 CGGTGGCAGTGGAGCGCGTGGTGGAGGCTCGAGGTGCGTGTGGCGCGCGCTGG 671  
Qy 137 GlnGlyLysArgHisAlaGlnAlaTrpAlaPheGluArg-----SerAspAlaAspVal 154  
Db 672 GCGGCAAGCGAGGTGATGTACACAGCTTCAAGGCGCTCGAGATTCGGTGACTAC 731  
Qy 155 PheLeuThrValAspSerAspThrTyrIleTyrProAspAlaLeuGlnGluLeuLys 174  
Db 732 GTGAGGTCTGTGACTCGGACACAGGTTGGACCCCATGCGACTGCTGGAGCTCGTGG 791  
Qy 175 ThrPheAsn---AspProThrValPheAlaAlaThrGlyHisLeuAsnValArgAsnArg 193  
Db 792 GTACTGGAGGAGACCCCGGTGAGGCTGTGTGGGAGCGTGGAGCTTAAACCT 851  
Qy 194 GlnThrAsnLeuLeuThrArgLeuThrAspIleArgTyrAspAsnAlaPheGlyValGlu 213  
Db 852 CTGAGCTCTCGGTGAGTCTCAAGCAGCTGCGTACTGCGTCAATGTGGAG 911  
Qy 214 ArgAlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGlyProLeuSerValTyr 233  
Db 912 CGGCTGTGACAGTACTTCCACTGTGTATCTGTCATCAGCGGTGCTTACGCGCTATAT 971  
Qy 234 ArgArgGluValValProAsnIleAspArgTyrIleAsnGlnThrPheLeuGlyLe 253  
Db 972 AGGAATAACCTCTTCAGCAGTCTTGTAGGCGGTGTACACACAGAGTCTCGGTACC 1031  
Qy 254 ProValSerIleGlyAspAspArgCysLeuThrAsnTyrAlaThrAspLeuGly---Lys 272  
Db 1032 CACGTACTTTTGGGATGACCGGACCTCACCAACCGCATGCTCAGCATGGGTATGCT 1091  
Qy 273 ThrValTyrGlnSerThrAlaLysCysIleThrAspValProAspLysMetSerThrTyr 292  
Db 1092 ACCAAGTACACTCCAGTCCGCTGCTACTACAGACGCGCTGCTCTTCTCGGTGG 1151  
Qy 293 LeuLysGlnGlnAsnArgTyrAsnLysSerPheArgGluSerIle----- 308  
Db 1152 CTGAGCAGCAGACACGCTGCTCAAGTGTACTTCCGTGAGTGGCTGTACACGCGCTC 1211  
Qy 309 -----IleSerValLysLysIleMetAsnAsn-----Pro 318  
Db 1212 TGGTGGCAGCGCACCATGCGTGGATGACCTACAGCGGTGTGCTCCGCGCTGTTCCC 1271  
Qy 319 PheValAlaLeuThrThrIleLeuGluValSerMetPheMetLeuValTyrSerVal 338  
Db 1272 TTCTCTGTGGCGGCACTGTGCTGCTG----- 1301  
Qy 339 ValAspPhePheValCysValAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeu 358  
Db 1302 -----TTCTACGCGCGCGCTGGCGCTGCTGGGTG-----CTG 1340  
Qy 359 ValIleIlePheIleValAlaLeuCysArg---AsnIleHisTyrMetLeuLysHisPro 377  
Db 1341 CTGTGGCTGGAGGCGTGGCACTGGCAGGCGGCTTCGCGGCTGCTGGCGGCGTGC 1400

Qy 378 LeuSerPheLeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeu 397  
Db 1401 CTGCGCATGGTCTTCTGTGCTCTACGCGCCCTCTACATGTGTGGCTCTGCTGCC 1460  
Qy 398 LysLeuTyrSerLeuPheThrIleArgAsnAlaAspTrpGlyThr-----ArgLysLys 415  
Db 1461 AAGTTCTCTGGCGCTAGTACCATGAACAGAGTGTGTGGGCACTCGCGCGGGAAG 1520  
Qy 416 Leu 416  
Db 1521 CTG 1523  
RESULT 7  
US-08-867-030B-5  
; Sequence 5, Application US/08867030B  
; Patent No. 5948900  
; GENERAL INFORMATION:  
; APPLICANT: Yother et al.  
; TITLE OF INVENTION: Streptococcus pneumoniae  
; TITLE OF INVENTION: Capsular Polysaccharide Genes and Flanking Regions  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Benjamin Aaron Adler, Ph.D., J.D.  
; STREET: 8011 Candle Lane  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch floppy  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh  
; SOFTWARE: Microsoft Word for Macintosh  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/867,030B  
; FILING DATE: June 2, 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/243,546  
; FILING DATE: May 16, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benjamin Aaron Adler, Ph.D., J.D.  
; REGISTRATION NUMBER: 35,423  
; REFERENCE/DOCKET NUMBER: D5923  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (713) 777-2321  
; TELEFAX: (713) 777-6908  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4951 bp  
; TYPE: nucleic acid  
; STRANDEDNESS: single-stranded  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
; DESCRIPTION: other nucleic acid  
; HYPOTHETICAL: no  
; ANTI-SENSE: no  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; IMMEDIATE SOURCE:  
; POSITION IN GENOME:  
US-08-867-030B-5  
Alignment Scores:  
Pred. No.: 3,76e-19 Length: 4951  
Score: 242.50 Matches: 114  
Percent Similarity: 43.84% Conservative: 78  
Best Local Similarity: 26.03% Mismatches: 183  
Query Match: 11.31% Indels: 65  
DB: 2 Gaps: 19  
US-09-469-200D-2 (1-417) x US-08-867-030B-5 (1-4951)

QY	16	pHeTrpValLeuLeuIleTy <sup>r</sup> ValAsnVal <sup>r</sup> TyrLeuPheGlyAlaLysGlySerLeuSer	35
DB	1389	TTTTTAIGRA---TACATTTATTTTAATGTTGTGGATTTTTCAGACATCATGATTTTCA	1345
QY	36	IleTyGlyPheLeuLeuIleAlaTy <sup>r</sup> LeuLeuLysMetSerLeuSerPhePhe---	54
DB	1346	T--TTCATTATGTTGTTTGTCTTTATCTTATCTTCTGTTGGCGGTATATATTTTCAT	1403
QY	55	-----TyrLysProPheLysGlyArgAlaGlyGlnTyrLysVal-----AlaAla	69
DB	1404	GCTGTCAGATATAGTCCTCAGCTGTGTAGTGTAGTCAGATGAGAGTTTATTTAGTCTCTGA	1463
QY	70	IleIleProSerTyrAsndLuAspAlaGluSerLeuLeuGluThrLeuLysSerValGln	89
DB	1464	ATTATCCCTGTCGTGATGAACCACTTAATCTTTTGAAGTGTACTGAATAGAAATTC	1523
QY	90	GlnGlnThrTyrProLeuAlaGluIleTyrValValAspAspGlySerAlaAspGluThr	109
DB	1524	AGA-----CATAAACCATCCGAAATTTATGTGGTTATTAAAGGCCCAAAACAGAGAGA	1577
QY	110	GlyIleLysArgIleGluAspTyrValArgAspThrGlyAspLeuSerSerAsnValIle	129
DB	1578	CTGTGAACACTTGTCTCATGATTT-----	1601
QY	130	ValHisArgSerGluLysAsnGln-----Gly	138
DB	1602	AATGAAAATTTAGAAAATATATGATCCCAATCAATGTTATTACACTCTCTGTCTCTGGC	1661
QY	139	LysArgHisAlaGlnAlaIleTrpAlaPheGluArg-----SerAspAlaAspValPheLeu	156
DB	1662	AAGAGAAATGCTATCGCGTTGGCTGGAGCATGTGGATTCGCAGAGTGATATTACAGTT	1721
QY	157	ThrValAspSerAspThrTyrIleTyrProAspAlaLeuGluLeuLeuLysThrPhe	176
DB	1722	CTAGTAGATAGTACATGATGAGCGCTAGAACCTTTGAGTGAGTGTCTGAGAGCTTTT	1781
QY	177	---AsnAspProThrValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThr	195
DB	1782	GTTTGGCATAAAAAATAGTGGGGTAAACGACAGACAAAAAATCTTGACCTGAGCGT	1841
QY	196	AsnLeuLeuThrArg-----LeuThrAspIleArgTyrAspAsnAlaPheGly	211
DB	1842	ATATCGTGCACATTTTGCTPAACCTTTAGAGGAAATTAGGCGAGAGGAACATATG---	1898
QY	212	ValGluArgAlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGlyProLeuSer	231
DB	1899	-----AAAGCATGNGTGTACTGGTAAAGTA---GGTGCTTACCTGGTGTGACAA	1946
QY	232	ValTyrArgArgGluValValProAsnIleAspArgTyrIleAsnGlnThrPheLeu	251
DB	1947	ATTGCTTTTAGAAATATAGTGGA---GAGAGTGTATACAAAGTTTATAAGAGACTTTTCATG	2005
QY	252	GlyIleProValSerIleGlyAspAspArgCys---LeuThrAsnTyrAlaThrAspLeuGln	271
DB	2006	GGATTTCATAGGAGACTTCTGATGATAGACGCTTACAAATTTTGACTTTTAAAAAAGG	2065
QY	271	y---LysThrValTyrGlnSerThrAlaLysCysIleThrAspValProAspLysMetSe	290
DB	2066	CTATAAACATGTTATGCAGAGATATCTCTCTGTGTATACAGATGCTCTACAAAGTGGAA	2125
QY	290	rThrTyrLeuLysGlnAsnArgTrpAsnLys---SerPhePheArgGluSerIleI	309
DB	2126	AAAGTTTCATTAGACACACACTAGGTGGCAGAGGTTCTCAGTATACAAATCTAAAGAT	2185
QY	309	eSerValLysLysIleMetAsnAsnProPheValAlaLeuTrpThrIleLeuGluValIse	329
DB	2186	GACTCCTTGGATGATTAGAAATGCCCTCTTATGTTTTTATTTATTATTACAGATATGAT	2245
QY	329	rMetPheMetMetLeuValTyrSerValValAspPhePheValGlyAsnValArgGluPh	349
DB	2246	TTTACCTATGCTACTTATAGCTTGTGGTGAATATATCTGTTGAAATATTTAAATAT	2305

```

Qy 349 e-----AspTrpLeuArgValLeuAlaPheLeuValIleIlePhe 362
Db 2306 AACTACAATCTTTATACAGCTTCATCGTGGAAATATTATTTATATGTTCTTTTGGAAAT 2365
Qy 362 eileValAlaLeu---CysArgAsnIleHisTyrMet-----LeuLysHisProLeuSe 379
Db 2366 GATTTTACGTTGGAGGAAGAACTTTAAAGCTATGCTAGAAATGAAGTGGTATTATGCT 2425
Qy 379 rPheLeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysIle 399
Db 2426 ATTCTTATTCCTGTTTATTAATACGTTTGTAGIATA---ATTATGTCCTCATTAGGCT 2482
Qy 399 uTyrSerLeuPheThrIleArgAsnAlaAsp-----TrpGlyThrArg 413
Db 2483 ATTAGACTT-----ATGAGATCTTCTGATGATTAGGGTGGGAAGTGG 2528

RESULT 8
PCT-US95-06119-5
; Sequence 5, Application PC/TUS9506119
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE CAPSULAR
; TITLE OF INVENTION: POLYSACCHARIDE GENES AND FLANKING REGIONS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06119
; FILING DATE: CONCURRENTLY HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/243,546
; FILING DATE: 16-MAY-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: AMCY018P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4951 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
PCT-US95-06119-5

Alignment Scores:
Pred. No.: 3,76e-19 Length: 4951
Score: 242.50 Matches: 114
Percent Similarity: 43.84% Conservative: 78
Best Local Similarity: 26.03% Mismatches: 183
Query Match: 11.31% Indels: 65
DB: 5 Gaps: 19

US-09-469-200D-2 (1-417) x PCT-US95-06119-5 (1-4951)

```



Query Match: 11.17% Indels: 101  
 DB: 4 Gaps: 19  
 US-09-469-200D-2 (1-417) x US-09-107-532A-2775 (1-1326)

QY 34 LeuSerIleTyrGlyPheLeuLeuLeuLeuLeuLeuVal----- 47  
 DB 67 TTGCTATTATGTTATTTTCTCGTTTCTATCCGATTTTAGCGGATTTGATGGTTT 126  
 QY 48 -----LysMetSerLeuSerPhePheTyrIlys----- 56  
 DB 127 ATGGTGCTGGTGTATGTTATTTTATACAAACAFAACAAAGAAGATGGGTAGATGTT 186  
 QY 57 -----ProPheLysGlyArgAlaGlyLysValAlaLalLeuLeu 71  
 DB 187 CCCTTGAGTGTGCAACCGTTC-----ATCCAGATCAATGGTG 222  
 QY 72 ProSerTyrAsnGluAsp-----AlaGluSerLeuLeuGluThrLeuLysSerValGln 89  
 DB 223 CCAGCCCAATATGAAGAAATGTGATCGATATACATATACATGAGTATTG-----ATGACG 276  
 QY 90 GlnGlnThrTyrProLeuAlaGluIleTyrValValAspAspGlySerAlaAspGluThr 109  
 DB 277 AAATTCAATTATCACAACACTAGGTATTAGTACAGACAGATGGTTCAACGGACCAACA 336  
 QY 110 Gly-----IleLysArgIleGluAspTyrValArgAspThrGlyAspLeuSerSer 126  
 DB 337 CCAGAAATATTAGTACTGATGATGAAAAATAC-----GCT 372  
 QY 127 AsnValIleValHisArgSerGluLysAsnGlnGlyLysArgHisAlaGlnAlaTrpAla 146  
 DB 373 AATCTCTGTGTGTTGCGATCGAAAAATAAAGAAAGCCCGGTTAATATCGGA 432  
 QY 147 PheGluArgSerAspAlaAspValPheLeuThrValAspSerAspThrTyrIleTyrPro 166  
 DB 433 CTGCGCTTTGCAAAAGGAAATGATTTAAGCAATGATGCAGACAGGTACCGGAGCCA 492  
 QY 167 AspAlaLeuGluGluLeuLeuLysThrPheAsnAspPro-----ThrValPheAla 183  
 DB 493 GATGCACATGATTAGGTATGTAATTTATTCATTCGCTGTGCAAGACATATGTCGCG 552  
 QY 184 AlaThrGlyHisLeuAsnValArgAsnGluThrAsnLeuLeuThrArgLeuThrAsp 203  
 DB 553 GTCATGCAATATGATGTCMAATCGT---ACCAATGATGTCMAATCTCAACG 609  
 QY 204 IleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSer---ValThrGlyAsn 222  
 DB 610 GTTGAATTCCTCAAGTATCGGGCATCATTAACGAACACACAGTCAGCTGTTTCGGTGGT 669  
 QY 223 IleLeuValCysSerGlyProLeuSerValTyrArgArgGluValValPro----- 240  
 DB 670 CTTTATGCGTACAGTGGAGCAATACGATGTATCGAAAGAAAGCCTTGATCGATGTGGCG 729  
 QY 241 -----AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGly 258  
 DB 730 GGATTCAGACAAAGATCGAGCAGACAGAC-----ATCAGTATGCT 771  
 QY 259 AspAspArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThr 278  
 DB 772 TGGATCATCAGTAAATGACTGTGTTA-----TCTGTTTTCCTCTCGAG 816  
 QY 279 AlaLysCysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArg 298  
 DB 817 ATATCTTTTTCATGGAATACCTGTGACGTTTGAAGATGCTATATATCAGACAGAAAGCT 876  
 QY 299 TrpAsnLysSerPhePheArgGluSerIleSerValLysLysIleMetAsnAsnPro 318  
 DB 877 TGGCAAAAGGTGGACACAGACTCGGCTGACGAATTTCAAAAAGGTGATGCTCCATCCT 936  
 QY 319 PheValAlaLeuTrp---ThrIleLeuGluValSerMetPheMetMetLeuValTyrSer 337  
 DB 937 TTATAACATATTGGAAGAACGATCATCTTCATTGATCAACATTTGAGTATTGCTGGTCA 996

QY 338 ValValAspPhePheVal----- 343  
 DB 997 ATA-----TTCTTTTGTATTTCGGTAGTTCTTTTGTGGATTAATCGGCAATTATGTA 1050  
 QY 344 -----GlyAsnValArgGluPheAspTyrPheArgValLeuAlaPheLeuValIleLeu 361  
 DB 1051 TATCAGGAACACTACGAACAGATTATATTACATTCAGTTTGTGTTCATCTGC 1110  
 QY 362 Phe-----IleValAlaLeuCys----- 367  
 DB 1111 TTTGAATGTGCGCAGGATTTTCCAGTTTCCTAGCTTCGCTGATTCTAGATGATCGACA 1170  
 QY 368 ArgAsnIleHieTyrMetLeuLysHisProLeuSerPheLeuLeuSerProPheTyrGly 387  
 DB 1171 AGAAATGTGAATACCTTCCTTTTCGCACCACTATATATGCTTTG-----TTCTGG 1221  
 QY 388 ValLeuHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThrIleArgAsn 407  
 DB 1222 ATATGACGCGCATAAACGATCGTGACGACTTTCATTCACGACTCAAAACGATTTAGGT 1281  
 QY 408 AlaAspTyrGlyThrArgLys 414  
 DB 1282 TATGGAAGCGGACATGCAAA 1302

RESULT 10  
 US-09-328-352-2963  
 ; Sequence 2963, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTE  
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 2963  
 ; LENGTH: 1260  
 ; TYPE: DNA  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-2963

Alignment Scores:  
 Pred. No.: 1,06e-19 Length: 1260  
 Score: 239.00 Matches: 101  
 Percent Similarity: 41.49% Conservative: 72  
 Best Local Similarity: 24.22% Mismatches: 168  
 Query Match: 11.14% Indels: 76  
 DB: 4 Gaps: 17

US-09-469-200D-2 (1-417) x US-09-328-352-2963 (1-1260)

QY 40 LeuLeuIleAlaTyrLeuLeuValLysMetSerLeuSerPhePheTyrLys----- 56  
 DB 61 TTGTGATGCTGTGACATGATGTAGGAGTCTCTGGTTTCTTTTAAAGGGAATAT 120  
 QY 57 -----ProPheLysGlyArgAlaGlyClnTyrLysValAlaIleLeu 71  
 DB 121 CATGACGACCAACTCCCTGAGCCATCAGTGAAGG-----TGCAATATTATT 171  
 QY 72 ProSerTyrAsnGluAspAlaGluSerLeuLeuGluThrLeuLysSerValGlnGln 91  
 DB 172 CCTTGCTTTTAAATGAAGAGCTAG---GTCCGCAACCAACGATTCGTATGCTTACAGACA 228  
 QY 92 ThrTyrProLeuAlaGluIleTyrValValAspAspLysSerAlaAspGluThrGlyIle 111  
 DB 229 AAATATCCGAATTTTGAAGTCATTCGTAAATCAGCGAAGTAGTGACAGCACTGCA--- 285  
 QY 112 LysArgIleGluAspTyrValArgAspThrGlyAspLeuSerSerAsnValIleValHis 131  
 DB 286 GAAATCTTTGATGATGCGCAGCTCAAGACGCCGATTAAGA-----GTTGTGCAC 336  
 QY 132 ArgSerGluLysAsnGlnGlyLysArgHisAlaGlnAlaTrpAlaPheGluArgSerAsp 151

```
Db 337 CTGCGAGAA---AACCAAGGTAAGGCGAGTCTTAAGATCGGAGTCTTAGTAAGTAAA 393
QY 152 AlaSpValPheLeuThrValAspSerAspThrTyrIleTyrProAspAlaLeuGluGlu 171
Db 394 TAGTAATCTAGTGTATTTAGCGAGACGCTTTATATACATCCTCATGCACTACTTGG 453
QY 172 LeuLeuLysThrPhe---AsnAspProThrValPheAlaAlaThrGlyHisLeuAsnVal 190
Db 454 CTATACAGCCTTTTAAATTTCCAGAATTTGGTCTGCTAACTGGGAATCCTAGAAAT 513
QY 191 ArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIleArgTyrAspAsnAlaPhe 210
Db 514 TTAATTCGCG---TCAAGTATTTAGTAAATTAACAAGTTGGCGAGTCTCTCTCAATATT 570
QY 211 GlyValGluArgAlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGlyProLeu 230
Db 571 GGTGTTAGTACCGCGACACAGAACATATGCGCGTATTTTACAGTATCAGGTGAAT 630
QY 231 SerValTyrArgArgGluValValValProAsnIleAspArgTyrIleAsnGlnThrPhe 250
Db 631 GCAGCATTTAGAAAGACCGCTTGGTT----- 657
QY 251 LeuGlyIleProValSerIleGlyAspAspArgCysLeuThr----- 264
Db 658 -----CGCGTGGCGCTTTGGTCTGATGACAAATATACGGAAGATATTGATCTCT 708
QY 265 ---AsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLysCysIleThr 283
Db 709 TGGAGCTCCAAATGGACCATTTGGACATTCATATATTCCTCAAGCACCTTTGTTATATC 768
QY 284 AspValProAspLysMetSerThrTyrLeuLysGlnAsnArgTyrAsnLysSerPhe 303
Db 769 TATATCGCGGAACCTTTAAAGCGCTTTTGGAAACAGCGTTTACGCTGGCACAGGTGG 828
QY 304 PheArgGluSerIleIleSerValLysLysIleMetAsnAsnProPheValAlaLeuIrp 323
Db 829 GTAGAGGTATTATTAGATATATCCGAAATGTTCAAACTCCGCTTGCCTGATGTGG 888
QY 324 ThrIleLeu-----GluValSerMetPheMetMetLeu 334
Db 889 CCGGTAAATGCTCGAAGCTTAAATCAGTATTATTGTCATATGTCATGATGATGATTTT 948
QY 335 ValTyrSerValValAspPheValGlyAsnValArgGluPhe----- 349
Db 949 ATTCTATTTTTGTGGGATTTGTTGCAATTTACCGCAACAATTTCAAATTAATTCAC 1008
QY 350 -----AspTyrLeuArgValLeu-----AlaPheLeuValIle 360
Db 1009 ATGCCGCGAGTGTATGCGGTGATATAGTGGGAACATGCTTGTTCATTTCTAGTAAGT 1068
QY 361 IlePheIle-----ValAlaLeuCysArgAsnIleHisTyrMetLeu 374
Db 1069 TTATGGATAGACCATGTTATGATCGTGGTGGCTTTTGTAGAAATTTATTTGGTAAAT 1128
QY 375 LysHisProLeuSerPhe---LeuLeuSerProPheTyrGlyValLeuHisLeuPheVal 393
Db 1129 TGGTATCATCTATTTTTCGTTACTGACGTTATTTACCACTGTT-----GTT 1176
QY 394 LeuGlnProLeuLysLeuTyrSerLeuPheThrIleArgAsnAlaAspTyr 410
Db 1177 GCAGTACCCAAACTATATTATAC-----ACCAAAACAGTGTCTGCTGG 1221
```

## RESULT 11

```
US-09-134-001C-1214
; Sequence 1214, Application US/09134.001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
```

```
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1214
; LENGTH: 1257
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1214

Alignment Scores:
Pred. No.: 3,82e-19 Length: 1257
Score: 234.50 Matches: 89
Percent Similarity: 47.09% Conservative: 97
Best Local Similarity: 22.53% Mismatches: 152
Query Match: 10.93% Indels: 57
DB: 4 Gaps: 16

US-09-469-200D-2 (1-417) x US-09-134-001C-1214 (1-1257)

QY 34 LeuSerIleTyrGlyPheLeuLeuIleAlaTyrLeuLeuValLysMet----- 49
Db 19 ATGCATCTGATTTAACTTTTCTATCCCAATTTTATGTCATTTTACTGGATAGTA 78
QY 50 ---SerLeuSerPhePheTyr-----LysProPhe-----Lys 59
Db 79 GGATCGATTTACTATTATTTTATTAAAGAAAACCTTTAATGATCATCTGTTAGTAAAA 138
QY 60 GlyArgAlaGlyGluTyrLys---ValAlaAlaIleIleProSerTyrAsnGluAspAla 78
Db 139 TCTCAACATCAACAGTTGAGGCACTCTCTTTTATTTAGTCTGCTACATGAA---AGT 195
QY 79 GluSerLeuLeuGluThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIle 98
Db 196 GAAACAGTTCAAGACACGCTTCTCTAGTATTTTATCTCTAGATATATCTGTAAGAAAT 255
QY 99 TyrValValAspAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrVal 118
Db 256 ATCATATATCAATGATGGAAGTCTCTGATAATCTGCT---GAAATCATCTATGACTTCAAG 312
QY 119 ArgAspThrGly-----AspLeuSerSerAsnValIleValHisArgSerGlu 134
Db 313 AAAATCATCATGATTTAAATTTGTTGACCTC-----GAA 345
QY 135 LysAsnGlnGlyArgHisAlaGlnAlaIlePheGluArgSerAspAlaAspVal 154
Db 346 GTCATAGAGGTAAGCTTAATGCACTCAATGAGGGAATCAACAGCATCTTACGAATAT 405
QY 155 PheLeuThrValAspSerAspThrTyrIleTyrProAspAlaLeuGluLeuLeuLys 174
Db 406 GTATGTGTTTACATGCTGACACTGTCTATGATGACGATCGCTTTTATATGATGATGAA 465
QY 175 ThrPhe---AsnAspProThrValPheAlaAlaThrGlyHisLeuAsnValArgAsnArg 193
Db 466 GACATTTAAAAAGATCCAAAATTTAGCGCAGCTACAGGTAATCCAGTATTCGTATATAA 525
QY 194 GlnThrAsnLeuLeuThrArgLeuThrAspIleArgTyrAspAsnAlaPheGlyValGlu 213
Db 526 ---AGTCTATTATTTAGGAAAAATACAGACCATGAAATATGCAAGTATTATTGGTGTATC 582
QY 214 ArgAlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGlyProLeuSerValTyr 233
Db 583 AAGCAAGTCAATCTCTTGCAGGCAATCAATACTATTTCAGGTGTTTTCACACTATT 642
QY 234 ArgArgGluValValValProAsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIle 253
Db 643 AAAAAGTGCACTC-----AAGATGATGATGTTTGGGATCTGACATGATCTAG 696
QY 254 ProValSerIleGlyAspAspArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThr 273
Db 697 GATATTGCTGTTTCATGAAACCTCATCTCTTTTGTATTAC-----GAAAT 741
```

Qy	274	Val	tyr	gln	ser	thr	ala	lys	cys	ile	thr	asp	val	pro	asp	lys	met	ser	thr	tyr	leu	293	
Db	742	AAG	TAC	GAA	ACC	ACG	CGT	CTC	TAT	GTG	TAT	GTC	GCT	GAA	ACT	AT	AGG	TGC	TTT	ATG	CG	801	
Qy	294	Lys	ser	gln	asn	arg	trp	-----	asn	lys	ser	phe	arg	lys	ile	leu	309						
Db	802	AAC	AA	GAG	GGT	CG	TGG	GCT	CA	AGG	GCG	CAT	GA	AGT	TACT	TTT	TA	AG	A	G	A	CG	861
Qy	310	Ser	val	lys	lys	ile	met	asn	pro	phe	val	ala	leu	thr	thr	ile	gln	val	ser	329			
Db	862	ACA	ATT	AAA	ACA	TGA	AAA	TAT	TCA	TAT	ATAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	921	
Qy	330	Met	phe	met	met	leu	val	tyr	ser	val	val	asp	phe	phe	val	-----	gly	asn	val	arg	347		
Db	922	ACA	TGG	GCT	AC	AT	CGT	ACT	AGT	TAT	TAT	TCT	TTT	TAT	TAT	TCT	TTT	TAT	TAT	TAT	TAT	981	
Qy	348	Glu	phe	asp	trp	leu	arg	-----	-----	-----	-----	-----	-----	-----	-----	-----	val	leu	ala	356			
Db	982	GAT	TAC	ACA	TAT	TAA	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	1041	
Qy	357	phe	leu	val	lle	lle	---phe	ile	val	ala	leu	cys	arg	asn	ile	his	tyr	met	leu	lys	375		
Db	1042	TTT	AT	CA	AT	AT	AT	CA	TTT	AC	AGT	TTC	TTT	TTT	AT	TAT	TAT	TAT	TAT	TAT	TAT	1101	
Qy	376	His	pro	leu	ser	phe	leu	leu	ser	pro	phe	tyr	gly	val	leu	his	390						
Db	1102	ATA	TAT	TGG	CGT	GAT	AT	TTT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	TAT	1146	

RESULT 12

```

US-09-328-352-1781
; Sequence 1781, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1781
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1781

```

Alignment Scores:	
Pred. No.:	1.86e-18
Score:	229.00
Percent Similarity:	44.13%
Best Local Similarity:	23.50%
Query Match:	10.68%
DB:	4
Length:	1263
Matches:	90
Conservative:	79
Mismatches:	156
Indels:	58
Gaps:	14

US-09-469-200D-2 (1-417) x US-09-328-352-1781 (1-1263)

QY	24	AsnValTyrLeuPheGlyAlaLysGlySerLeuSerIle-----TyrGlyPhe	39
Db	1	AACATGACATATCTTGCTGCTATGTCATGCTTAAATTCGTTTCTATTATCCA	60
QY	40	LeuLeuIleAlaTyrLeuValLysMetSerLeuSerPhePheTyrLysProPheLys	59
Db	61	TTATTTATGTCGTATTATTGGATGATGGCGGCATCTTTCTATGCAAG-----GAA	114
QY	60	GlyArgAlaGlyGlnTyr-----LysValAlaIle	69
Db	115	CGTAAAGCTCGCGCTTATCATGCCAGCTCCATTCAAAAGCTATCTTAAGTGTGCTGTG	174
QY	70	IleIleProSerTyrAsnGluAspAlaGluSerLeuGluThrLeuLysSerValGln	89
Db	175	CTATTACTCTGTTTATTGAA---GGGGATATGACAGCAACGATTATCTATGCTTTA	231

QY	90	GlnGlnThrThrProLeuAlaGluLeuTyrValValAspAspGlySerAlaAspGluThr	129
Db	232	AAATGGATATTCCTATTTCGAAGCATTTGCGATTAAATGATGCGATTCGGATAATACA	291
QY	110	GlyIleLysAlaGlieGluAspTyrValArgAspThrGlyAspLeuSerSerAsnValIle	129
Db	292	GGT-----GAAGTGGTGGATCGGTAGCTGAACACATGAAANAACCTCGT-----GTT	339
QY	130	ValHisArgSerGlnLysAsnGlnGlyLysArgHisAlaGlnAlaTrpAlaPheGluArg	149
Db	340	GTTTCATCTTGCACAA--AACCAAGGTAAAGCATGGGGTTGCACCTGGAAGCTTAATG	396
QY	150	SerAspAlaAspValPheLeuThrValAspSerAspThrTyrIleTyrProAspAlaLeu	169
Db	397	ACCGATGCCGAGTTTAAATGGTATGATGTGTGATGCTTTACTTACCCGCATGCACCA	456
QY	170	GluGluLeuLeuLysThrPhe---AsnAspProThrValPheAlaAlaThrGlyHisLeu	188
Db	457	AGTGGATGGTTCGTATTTCTCAAAATCCGAGCTGTGCAGCGCTGACGGGTAAACCG	516
QY	189	AsnValArgAsnArgGlnThrAsnLeuLeuThrAlaLeuThrAspIleArgTyrAspAsn	208
Db	517	CGTATCCGTCGCTCAACT---TTATTAGTCTGATTCAGTTCGTGGATGTTCTTCGTG	573
QY	209	AlaPheGlyValGluArgAlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGly	228
Db	574	ATTGTCCGCATGATTAAACGTGCTCAGCGTACCTTTGSCCGTTTATTACCGTATCTGGT	633
QY	229	ProLeuSerValTyrArgArgGluValValProAsnIleAspArgTyrIleAsnGln	248
Db	634	GTTATTACGGCTTTTCGTAAAGAGCGTGT-----CACCAA	669
QY	249	ThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThrAsnTyrAlaThr	268
Db	670	GTGGCATTTGGTCACCTAATATCTTGCACGAGACATCATATTACTTGGAAATTCGAA	729
QY	269	AspLeuGlyLysThrVal---TyrGlnSerThrAlaLysCysIleThrAspValProAsp	287
Db	730	CGTCCGGCTGGACATTCGTTTAAACCGAATGCTTGTGTTGGATTTTAATCCCTGAA	789
QY	288	LysMetSerThrTyrIleLysGlnGlnAsnArgTrpAsnLysSerPhePheArgGluSer	307
Db	790	ACTTTAAACGGTTTGGGAACAACCGTTTACGTTTGGCAATGGGTGGAGCACAGGTATTA	849
QY	308	IleIleSerValLysIleMetAsnAsnProPheValAlaLeuThrTrpIleLeuGlu	327
Db	850	ATTAATAATATGATGCTTCACTAACCAAAATGAATTTTATGGCCCTTAATG---906	
QY	328	ValSerMetPheMetMetLeuValTyrSerValValAspPhePheValGlyAsnValArg	347
Db	906	-----906	
QY	348	GluPheAspTrpLeuArgValIleAlaPheLeuValIleIlePheIleValAlaLeuCys	367
Db	907	---TTTGAGCTTTGCTTAACCTTAGTCTGGTCATATTATTATGCTGGCGATGGCTTTGTTG	963
QY	368	ArgAsnIleHisTyrMetLeuLysHisPro---LeuSerPheLeuLeuSerProPhe---	385
Db	964	TGGTTGGTACATTTATTTTACCGGTCCCTGCATACCGCTGTGAGGCTCTCCATCTTTTA	1023
QY	386	---TyrGly	387
Db	1024	CCGTATGGT	1032

## RESULT 13

```

RESOUR 13
US-08-961-527-86/c
; Sequence 86, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:

```

[illegible]

; SOFTWARE: PatentIn ver. 2.1

; SOFTWARE: PatentIn ver. 2.1

; SEQ ID NO 16  
; LENGTH: 4066  
; TYPE: DNA  
; ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-16

## Alignment Scores:

Pred. No.: 8,74e-10 Length: 4066  
Score: 166.00 Matches: 78  
Percent Similarity: 41.64% Conservatives: 64  
Best Local Similarity: 22.87% Mismatches: 122  
Query Match: 7.74% Indels: 77  
DB: 15 Gaps: 15

US-09-469-200D-2 (1-417) x US-09-499-203-16 (1-4066)

QY 21 IleTyrValAsnValTyrLeu-----PheGlyAlaLysGly-SerLeuSerIle-- 36  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3117 ATTACATCAATTACTTATCATGATTAGTTTGGTTCATCTGGTTCATTCGGTTG 3176  
QY 37 -----TyrGlyPheLeuLeuIleAlaTyrLeuValLysMetSe 50  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3177 TTGTGGAAATTTCCAGCTATATTCAATCTAGT-----TTGATAAAACATC 3227  
QY 50 rLeu-----SerPhePheTyrLysProPheLysGlyArgAlaGlyGlnTyrLysValAl 68  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3228 AAAAAAGCATCAAGTCATGCAAGAACCAITGAA-----TTACCAATGGTATC 3275  
QY 68 aAlaIleIleProSerTyrAsnGluAspAlaGluSerLeuLeuGluThrLysSerVa 88  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3276 TATTGTATTACCTATATATAAT---TCAGGACAAACATTATACCAGTGTATTCAGTCGAT 3332  
QY 88 lGlnGlnGlnThrTyrPro-----LeuAlaGluIleTyrValValAspGlySerAl 106  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3333 TAGTCATCACTCACTCAACAAACATTAATCCAAATATTCCCGTGAATCAAGCAC 3392  
QY 106 aAspGluThrGlyLeLysArgIleGluAspTyrValArgAspThrGlyAspLeuSerSe 126  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3393 TGAT-----AACAG 3401  
QY 126 rAsnValIleValHisArgSerGlu----- 134  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3402 TTTTACAGTGTATTAATCAGGTCAAGCTGATTTCCCAATACTTAGAATCAGTGGATGAA 3461  
QY 135 -LysAsnGlnGlyLysArgHisAlaGlnAlaIleTyrPheGluArgSerAspAlaAspVa 154  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3462 CACAGATCAAGTAAGCGCGTGCCTAAATGCTGCTATTATTAATAGTATGGGCAATA 3521  
QY 154 lPheLeuThrValAspSerAspThrTyrIleTyrProAspAlaLeuGluLeuLeuLy 174  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3522 TATTATTAAATTTGGTACTGATGTTGTTAGAACCAATGCCCTAAACCGGTTTGTGCT 3581  
QY 174 sThrPhe---AsnAspProThrValPheAlaIleThrGlyHisLeu----- 188  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3582 TTATTTGAAATCACTCAGAAATGTAGTACCACTGGTACGATTCGACAAACAAAAA 3641  
QY 189 -----AsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIleAr 205  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3642 AATGATTCAAAAACGCAAGTAATGGCTTAATGCTACAGTTAAACGAATATTGGA 3701  
QY 205 gTyrAspAsnAlaPheGlyValGluArgAlaIleGlnSerValThrGlyAsnIleLeuVa 225  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3702 GTATGCAATATCTTATCATCGCCGCTAGTATTGAAACCGGTGGAATCGTTGTTCCAC 3761  
QY 225 lCysSerGlyProLeuSerValTyrArgArgGluValValProAsnIleAspArgTy 245  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3762 ATGTACAGGTGCAATTTACGCGTTTACAGCTGATGTATTAGTT----- 3804  
QY 245 rIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThr-- 264  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3805 -----CAGACATTTATGTATATGTGTGACACGGTGTGTAAGATCACTCATATGACATT 3857  
QY 265 -AsnTyrAlaThrAspLeuGlyLysThrVal---TyrGlnSerThrAlaLysCysIleTh 283

Db 3858 TCAACTGCGATTCCGCTAGCAAGCGTATTGGTTCTGTGATGATGCCATGTTTATGT 3917  
QY 283 rAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTTPAsnLys---- 301  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3918 TGAACCAATATCAGTTATAGTCACTTATTATTACACGCGCAGCTTGGCAGCGTGGCA 3977  
QY 302 -----SerPhePheArgGluSerIleIleSerValLysLysIleMetAsnAs 317  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
3978 AATTGAAGTCGCACAAAAATTTCATGACAGATAAGTTGAGTGTCCGTCAGATTTTACTAA 4037  
QY 317 n 317  
Db 4038 C 4038  
RESULT 15  
US-09-499-203-1  
; Sequence 1, Application US/09499203  
; Patent No. 6570065  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMANN, Jens  
; APPLICANT: WELSH, Thomas  
; APPLICANT: QUANZ, Martin  
; APPLICANT: KNUTH, Karola  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
; FILE REFERENCE: 147-196P  
; CURRENT APPLICATION NUMBER: US/09/499,203  
; CURRENT FILING DATE: 2000-02-08  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 9321  
; TYPE: DNA  
; ORGANISM: Leuconostoc mesenteroides  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (678)...(6848)  
US-09-499-203-1

## Alignment Scores:

Pred. No.: 3,46e-09 Length: 9321  
Score: 166.00 Matches: 78  
Percent Similarity: 41.64% Conservatives: 64  
Best Local Similarity: 22.87% Mismatches: 122  
Query Match: 7.74% Indels: 77  
DB: 15 Gaps: 15

US-09-469-200D-2 (1-417) x US-09-499-203-1 (1-9321)

QY 21 IleTyrValAsnValTyrLeu-----PheGlyAlaLysGly-SerLeuSerIle-- 36  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
8372 ATTACATCAATTTACTTATCATGATTAGTTTGGTTCATCTGGTTCATTCGGTTG 8431  
QY 37 -----TyrGlyPheLeuLeuIleAlaTyrLeuLeuValLysMetSe 50  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
8432 TTGTGAAATTTCCAGCTATATTATTCATCTAGT-----TTGATAAAACATC 8482  
QY 50 rLeu-----SerPhePheTyrLysProPheLysGlyArgAlaGlyGlnTyrLysValAl 68  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
8483 AAAAAAGCATCAAGTCATGCAAGAACCAITGAAA-----TTACCAATGGTATC 8530  
QY 68 aAlaIleIleProSerTyrAsnGluAspAlaGluSerLeuLeuGluThrLysSerVa 88  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
8531 TATTGTATTACCTATATATAAT---TCAGGACAAACATTATACCAGTGTATTCAGTCGAT 8587  
QY 88 lGlnGlnGlnThrTyrPro-----LeuAlaGluIleTyrValValAspGlySerAl 106  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
8588 TAGTCAATCAACCTATCCAAACAAATTAATCCAAATATTTCGCGTGAATAATCAAGCAC 8647  
QY 106 aAspGluThrGlyLysArgIleGluAspTyrValArgAspThrGlyAspLeuSerSe 126  
Db |||||:|||||: |||: |||||: |||: |||: |||||:  
8648 TGAT-----AACAG 8656

Search completed: October 2, 2003, 16:14:52  
Job time : 269 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 2, 2003, 12:56:43 ; Search time 334 Seconds  
(without alignments)  
3184.760 Million cell updates/sec

Title: US-09-469-200D-2

Perfect score: 2145

Sequence: 1 MRTLKLNITVAFSIFWVL.....KLYSLFTIRNADWGRKILL 417

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet\_p2n.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09469200/runat\_02102003\_090349\_22849/app\_query.fasta\_1.583  
-DB=Published Applications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOFCIL-0 -LOOPEXT-0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=ptc -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09469200.@cgn.1.1.221@runat\_02102003\_090349\_22849  
-NCPU=6 -ICPU=3 -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA:\*\*

1: /cgn2.6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2.6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2.6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2.6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2.6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2.6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2.6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2.6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2.6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2.6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2.6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2.6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2.6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2.6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2.6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2.6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2.6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2145	100.0	1254	10	US-09-879-959-1 Sequence 1, Appli

2	2145	100.0	1254	14	US-10-011-768B-1	Sequence 1, Appli
3	2145	100.0	1254	14	US-10-011-771B-1	Sequence 1, Appli
4	2145	100.0	1254	14	US-10-172-527-1	Sequence 1, Appli
5	2138	99.7	1251	12	US-10-326-185-1	Sequence 1, Appli
6	1976	92.1	5158	12	US-10-326-185-108	Sequence 108, App
7	1695.5	79.0	1251	12	US-10-326-185-102	Sequence 102, App
8	1695.5	79.0	3466	14	US-10-172-527-11	Sequence 11, Appl
9	1586.5	74.0	1257	12	US-10-326-185-92	Sequence 92, Appl
10	1586.5	74.0	1440	14	US-10-172-527-13	Sequence 13, Appl
11	1586.5	74.0	1512	14	US-10-124-222-1	Sequence 1, Appli
12	1586.5	74.0	1512	14	US-10-117-795-1	Sequence 1, Appli
13	977.5	45.6	1200	14	US-10-172-527-17	Sequence 17, Appl
14	463	21.6	536165	11	US-09-939-964-1	Sequence 1, Appli
15	440	20.5	1665	11	US-09-902-939-3	Sequence 3, Appli
16	439	20.5	1659	11	US-09-902-939-2	Sequence 2, Appli
17	439	20.5	4194	14	US-10-262-526-3	Sequence 3, Appli
18	433.5	20.2	3003	11	US-09-918-624B-64	Sequence 64, Appl
19	433.5	20.2	3003	14	US-10-262-526-1	Sequence 1, Appli
20	433.5	20.2	4018	12	US-09-814-353-19942	Sequence 19942, A
21	420.5	19.6	3387	14	US-10-084-817-184	Sequence 184, App
22	369	17.2	1317	14	US-10-156-761-4447	Sequence 4447, Ap
23	369	17.2	9025608	14	US-10-156-761-1	Sequence 1, Appli
24	367	17.1	1752	11	US-09-902-939-1	Sequence 1, Appli
25	366.5	17.1	2116	13	US-10-043-523-1	Sequence 1, Appli
26	343.5	16.0	1740	14	US-10-011-768B-7	Sequence 7, Appli
27	343.5	16.0	1740	14	US-10-011-771B-7	Sequence 7, Appli
28	336.5	15.7	1740	10	US-09-879-959-8	Sequence 8, Appli
29	336.5	15.7	1740	14	US-10-172-527-8	Sequence 8, Appli
30	245	11.4	1380	14	US-10-172-527-15	Sequence 15, Appl
31	242.5	11.3	7430	8	US-08-781-986A-260	Sequence 260, App
32	241.5	11.3	2370	14	US-10-156-761-4813	Sequence 4813, Ap
33	234.5	10.9	4500	10	US-09-771-003-1	Sequence 1, Appli
34	233.5	10.9	570	10	US-09-880-107-3449	Sequence 3449, Ap
35	213	9.9	2250	14	US-10-156-761-5099	Sequence 5099, Ap
36	195.5	9.1	13884	10	US-09-070-927A-341	Sequence 341, App
37	187.5	8.7	17276	10	US-09-870-759-83	Sequence 83, Appl
38	187.5	8.7	17276	12	US-09-751-708A-83	Sequence 83, Appl
39	184.5	8.6	3498	10	US-09-801-368-75	Sequence 75, Appl
40	175	8.2	987	12	US-09-769-744A-47	Sequence 47, Appl
41	173	8.1	1152	10	US-09-974-300-1802	Sequence 1802, Ap
42	171.5	8.0	3396	10	US-09-801-368-71	Sequence 71, Appl
43	169	7.9	2610	12	US-10-246-330-1	Sequence 1, Appli
44	164.5	7.7	490	11	US-09-918-995-27210	Sequence 27210, A
45	160	7.5	2892	10	US-09-801-368-73	Sequence 73, Appl

#### ALIGNMENTS

RESULT 1  
US-09-879-959-1  
; Sequence 1, Application US/09879959  
; Patent No. US20020160489A1  
; GENERAL INFORMATION:  
; APPLICANT: Weigel, Paul H  
; APPLICANT: Kumari, Kshama  
; APPLICANT: DeAngelis, Paul  
; TITLE OF INVENTION: STREPTOCOCCUS EQUISMILIS HYALURONAN SYNTHASE GENE AND EXPRESS  
; TITLE OF INVENTION: IN BACILLUS SUBTILIS  
; FILE REFERENCE: 3554.049  
; CURRENT APPLICATION NUMBER: US/09/879,959  
; PRIOR FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: 09/469,200  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 09/178,851  
; PRIOR FILING DATE: 1998-10-26  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1254  
; TYPE: DNA  
; ORGANISM: Streptococcus equismilis  
US-09-879-959-1

## Alignment Scores:

Pred. No.: 1.97e-238 Length: 1254  
 Score: 2145.00 Matches: 417  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0

US-09-469-200D-2 (1-417) x US-09-879-959-1 (1-1254)

QY 1 MetArgThrLeuLysAsnLeuThrValValAlaPheSerIlePheTrpValLeuLeu 20  
 DB 1 ATGAGAACATTAAACAACTCATACCTGTTGGCTTTAGTATTTTGGGTACTGTTG 60  
 QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40  
 DB 61 ATTACGTCATGTTTATCTCTTGGTCTTAAGAACTTGTCATTTATGGCTTTTGG 120  
 QY 41 LeuIleAlaTyrLeuLeuValLysMetSerLeuSerPheTyrLysProPheLysGly 60  
 DB 121 CTGATAGCTTACCTATTAGTCAAAATGCTTATGCTTTTACAAGCCATTAAAGGA 180  
 QY 61 ArgAlaGlyGlnTyrLysValAlaAlaIleProSerTyrAsnGluAspAlaGluSer 80  
 DB 181 AGGCTGGCAATATTAAGTTGCGACCTTATTCCTCTTATAACGAAGATCGTAGTCA 240  
 QY 81 LeuLeuGluThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVal 100  
 DB 241 TTGCTAGAGACCTTAAAGTGTTCAGCAGCAAACTATCCCTAGCAGAAATTTATGTT 300  
 QY 101 ValAspGlySerAlaAspGluThrGlyLysArgIleGluAspTyrValArgAsp 120  
 DB 301 GTTACAGTGAAGTGGCTGATGACAGAGTATTAAAGCGCATTAAGCACTATGCGTGAC 360  
 QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140  
 DB 361 ACTGGTCACTATCAGCAATGCTATGTTTCATCGCTCAGAGAAATCAAGAAAGCGT 420  
 QY 141 HisAlaGlnAlaTrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160  
 DB 421 CATGCAGAGCCTGGGCTTTGAAAGATCAGACGCTGATGCTTTTGGCCGTGACTCA 480  
 QY 161 AspThrTyrIleTyrProAspAlaLeuGluLeuLeuLysThrPheAsnAspProThr 180  
 DB 481 GATACCTTATATCCCTGATGCTTAGAGGAGTGTGTTAAACACCTTAAACCCCAACT 540  
 QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200  
 DB 541 GTTTTTCCTCGCAGGGTCACTTAAATGTCAGAAATAGACAAACCAATCTCTAACACGC 600  
 QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThr 220  
 DB 601 TTGACAGATATTCGCTATGATGATGCTTTTGGCTTGAACGAGCTGCCAATCCGTACA 660  
 QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgArgGluValValPro 240  
 DB 661 GGTATATCTGTTGCTCAGTCCGCTTAGCTTAGACGCGAGGTGTTGTCCT 720  
 QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260  
 DB 721 AACATAGATAGATACATCAACAGACCTCTCGGTATTCCTAGTAGTATGGTGATGAC 780  
 QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLys 280  
 DB 781 AGGTGCTTGACCACTATGCAACTGATTAGGAAGAGCTGTTTATCAATCCACTGCTAAA 840  
 QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsn 300  
 DB 841 TGTATTACAGATGTTCCGACAGATGCTACTTACTTACTTGAAGCAGCAAAACCGGTGAAC 900  
 QY 301 LysSerPhePheArgGluSerIleLeuSerValLysLysIleMetAsnAsnProPheVal 320  
 DB 901 AAGTCCCTCTCTAGACAGTCCATTTTCTGTTAAGAAATCATGAACAAATCCTTTTGTGTA 960

## RESULT 2

US-10-011-768B-1

; Sequence 1, Application US/10011768B  
 ; Publication No. US20030073221A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weigel, Paul H.  
 ; APPLICANT: DeAngelis, Paul  
 ; APPLICANT: Kumari, Kshama  
 ; TITLE OF INVENTION: Hyaluronan Synthase Gene and Uses Thereof  
 ; FILE REFERENCE: 3554.011  
 ; CURRENT APPLICATION NUMBER: US/10/011,768B  
 ; CURRENT FILING DATE: 2001-12-11  
 ; PRIOR APPLICATION NUMBER: US 09/178,851  
 ; PRIOR FILING DATE: 1998-10-26  
 ; PRIOR APPLICATION NUMBER: US 60/064,435  
 ; PRIOR FILING DATE: 1997-10-31  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 1254  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus equisimilis

US-10-011-768B-1

## Alignment Scores:

Pred. No.: 1.97e-238 Length: 1254  
 Score: 2145.00 Matches: 417  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 14 Gaps: 0

US-09-469-200D-2 (1-417) x US-10-011-768B-1 (1-1254)

QY 1 MetArgThrLeuLysAsnLeuThrValValAlaPheSerIlePheTrpValLeuLeu 20  
 DB 1 ATGAGAACATTAAACAACTCATACCTGTTGGCTTTAGTATTTTGGGTACTGTTG 60  
 QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40  
 DB 61 ATTACGTCATGTTTATCTCTTGGTCTTAAGAACTTGTCATTTATGGCTTTTGG 120  
 QY 41 LeuIleAlaTyrLeuLeuValLysMetSerLeuSerPheTyrLysProPheLysGly 60  
 DB 121 CTGATAGCTTACCTATTAGTCAAAATGCTTATGCTTTTACAAGCCATTAAAGGA 180  
 QY 61 ArgAlaGlyGlnTyrLysValAlaAlaIleProSerTyrAsnGluAspAlaGluSer 80  
 DB 181 AGGCTGGCAATATTAAGTTGCGACCTTATTCCTCTTATAACGAAGATCGTAGTCA 240  
 QY 81 LeuLeuGluThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVal 100

```
Db 241 TTGCTAGACAGCTTAAAGAGTGTTCAGCAGCAAAACCTATCCCTAGCAGAAATTTATGTT 300
QY 101 ValAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAsp 120
Db 301 GTTGACGATGAAGTGTGATGAGACAGGTATTAAAGCATTGAACACTATGTGCGTGAC 360
QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140
Db 361 ACTGGTGACCTATCAGCAATGTCATTGTTTCATCGGTGAGAGAAAATCAAGAAAGCGT 420
QY 141 HisAlaGlnAlaTrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160
Db 421 CATGCCAGCGCTGGCGCTTTGAAGATCAGACGCTGATGCTCTTTTGACCGTTGACTCA 480
QY 161 AspThrTyrIleTyrProAspAlaLeuGluLeuLysThrPheAsnAspProThr 180
Db 481 GATACCTATATCTACCTGATGCTTGAAGATCAGACGCTGATGCTCTTTTGACCGTTGACTCA 540
QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200
Db 541 GTTTTGTGCGCGGTGCTACCTTAATGTGAGAAATAGACAAACCAATCTCTTAACACGC 600
QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaGlnSerValThr 220
Db 601 TTGACAGATATTCGTATGATGATGCTTTTGGCGTTGAAGAGCTGCCCAATCCGTTACA 660
QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgArgGluValValPro 240
Db 661 GGTAAATACCTTGTGTTGCTCAGTCCGCTTAGCGTTTACAGACGCGAGGTGGTGTGCT 720
QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260
Db 721 AACATAGATAGATACATCAACACGACCTTCTCGGTGATTCCTGTAAGTATGTTGATGAC 780
QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLys 280
Db 781 AGTGTGTTGACCACTATGCACTGATTTAGGAAGAGCTGTTATATCAATCCACTGCTAAA 840
QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsn 300
Db 841 TGTATTACAGATGTTCTGACAGATGCTACTTACTTGTGAGCAAAACCCCTGGAAAC 900
QY 301 LysSerPheArgGluSerIleLysSerValLysLysIleMetAsnAsnProPheVal 320
Db 901 AAGTCTCTCTTTAGAGAGTCCATTTCTGTTAAGAAAATCATGACAAATCCTTTGTA 960
QY 321 AlaLeuTrpThrIleLeuGluValSerMetPheMetMetLeuValTyrSerValAsp 340
Db 961 GCCCTATGGACCATCTAGGTGCTATGTTTATGATGCTGTTTATCTGTTGTTGAT 1020
QY 341 PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle 360
Db 1021 TCTTTTGTAGCAATGTCAGAGAAATTCATTTGCTCAGGCTTGTAGCTTTCTGGTGAT 1080
QY 361 IlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPhe 380
Db 1081 ATCTTCATGTTGCTGCTGTGCGAATTCATTCATGATGCTGTTTATCTGTTGTTGAT 1140
QY 381 LeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyr 400
Db 1141 TTGTTATCTCGCTTTATGGGTGCTGATTTGTTGCTTACAGCCCTTGAATATAT 1200
QY 401 SerLeuPheThrIleArgAsnAlaAspTrpGlyThrArgLysLysLeuLeu 417
Db 1201 TCTCTTTTACTATTAGAAATGCTGACTGGGGAACAGTAAAAAATATTA 1251
```

## RESULT 3

US-10-011-771B-1

; Sequence 1, Application US/10011771B

; Publication No. US20030082780A1

; GENERAL INFORMATION:

; APPLICANT: Weigel, Paul H.

; APPLICANT: DeAngelis, Paul

```
; APPLICANT: Kumari, Kshama
; TITLE OF INVENTION: Hyaluronan Synthase Gene and Uses Thereof
; FILE REFERENCE: 3554.011
; CURRENT APPLICATION NUMBER: US/10/011.771B
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: US 09/178,851
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: US 60/064,435
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO. 1
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Streptococcus equisimilis
US-10-011-771B-1
```

## Alignment Scores:

Pred. No.:	1,97e-238	Length:	1254
Score:	2145.00	Matches:	417
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-09-469-200D-2 (1-417) x US-10-011-771B-1 (1-1254)

```
QY 1 MetArgThrLeuLysAsnLeuIleThrValValAlaPheSerIlePheTrpValLeuLeu 20
Db 1 ATGAGAACATTAATAAAACCTCATAACTGTTGGGCTTTAGTATTTTGGGTACTGTTG 60
QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40
Db 61 ATTACGCTCAATGTTTATCTCTTTGGTGTCTAAAGGAAGCTTGCAATTTATGGCTTTTG 120
QY 41 LeuIleAlaTyrLeuLeuValLysMetSerLeuSerPhePheTyrLysProPheLysGly 60
Db 121 CTGATACCTTACCTATTAGTCAAAATGCTCTTATCCCTTTTACAGCCATTTAAGGGA 180
QY 61 ArgAlaGlyGlnTyrLysValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSer 80
Db 181 AGGCTGGCAATATAAGGTTGCGAGCAGCATTTCCCTTTAAGAGAGTCTGAGTCA 240
QY 81 LeuLeuGluThrLeuLysSerValGlnGlnGlnThrTyrProLeuAlaGluIleTyrVal 100
Db 241 TTGCTAGAGACCTTAAAAGTGTTCAGAGCAAAACCTATCCCTAGCAGAAATTTATGTT 300
QY 101 ValAspAspLysSerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAsp 120
Db 301 GTTGACGATGAAGTGTGATGAGACAGGTATTAAAGCATTGAACACTATGTGCGTGAC 360
QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140
Db 361 ACTGGTGACCTATCAGCAATGTCATTGTTTCATCGGTGAGAGAAAATCAAGAAAGCGT 420
QY 141 HisAlaGlnAlaTrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160
Db 421 CATGCCAGCGCTGGCGCTTTGAAGATCAGACGCTGATGCTCTTTTGACCGTTGACTCA 480
QY 161 AspThrTyrIleTyrProAspAlaLeuGluLeuLysThrPheAsnAspProThr 180
Db 481 GATACCTATATCTACCTGATGCTTTTAGAGAGTGTGTTAAAACCTTTAAGACCAACT 540
QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200
Db 541 GTTTTGTGCGCGGTGCTACCTTAATGTGAGAAATAGACAAACCAATCTCTTAACACGC 600
QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaGlnSerValThr 220
Db 601 TTGACAGATATTCGTATGATGATGCTTTTGGCGTTGAAGAGCTGCCCAATCCGTTACA 660
QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgArgGluValValPro 240
```

Db 661 GGTAATATCCCTTTGCTCAGGTCCGCTTAGCGTTTACAGACGCGAGGTGGTGTCTCT 720  
 QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260  
 Db 721 ACATAGATAGATACATCAACAGACCTCTCTGGGTATTCCTGTAAGTATTTGGTGTATGAC 780  
 QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyIleThrValTyrGlnSerThrAlaLys 280  
 Db 781 AGTGCTTGACCAACTATCAACTGATTTAGGAAGACGTGTTATCAATCCACTGCTAAA 840  
 QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnAsnArgTrpAsn 300  
 Db 841 TGTATTACAGATGTTCTCTGACAGATGCTACTTACTTGAAGACGCAAAACCGCTGGAAC 900  
 QY 301 LysSerPhePheArgGluSerIleLeuValLysLysIleMetAsnAsnProPheVal 320  
 Db 901 AGTCTCTTTTAGAGAGTCCATATTCTGTTAGAAATCATGAACATCCCTTTTGTGA 960  
 QY 321 AlaLeuTrpThrIleLeuGluValSerMetPheMetMetLeuValTyrSerValValAsp 340  
 Db 961 GCCCTATGGACCACTCTGAGGTGCTATGTTATGATGCTGTTTATCTGTTGCTGGAT 1020  
 QY 341 PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle 360  
 Db 1021 TCTCTTTAGGCAATGTCAGAAATTTGATGGCTCAGGTTTATGCTTCTGCTGAT 1080  
 QY 361 IlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPhe 380  
 Db 1081 ATCTTCATTTGCTGCTGTGCGGAACATTCATACATGCTTAAGCACCCTGCTCTTC 1140  
 QY 381 LeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyr 400  
 Db 1141 TTGTTATCTCGCTTTATGGGTGCTGCTATTTGTTGCTACAGCCCTTGAATATAT 1200  
 QY 401 SerLeuPheThrIleArgAsnAlaAspTrpGlyThrArgLysLysLeuLeu 417  
 Db 1201 TCTCTTTTACTATTAGAAATGCTGACTGGGGAACACGTAAAAAATTATTA 1251

RESULT 4

US-10-172-527-1  
 ; Sequence 1, Application US/10172527  
 ; Publication No. US20030092118A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weigel, Paul H  
 ; APPLICANT: Kumari, Kshama  
 ; APPLICANT: DeAngelis, Paul  
 ; TITLE OF INVENTION: HYALURONAN SYNTHASE GENES AND EXPRESSION THEREOF IN BACILLUS SUBT  
 ; FILE REFERENCE: 3554.048  
 ; CURRENT APPLICATION NUMBER: US/10/172,527  
 ; PRIORITY FILING DATE: 2002-06-13  
 ; PRIOR APPLICATION NUMBER: 60/297,788  
 ; PRIOR FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: 60/297,744  
 ; PRIOR FILING DATE: 2001-06-13  
 ; PRIOR APPLICATION NUMBER: 09/469,200  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 09/178,851  
 ; PRIOR FILING DATE: 1998-10-26  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1  
 ; LENGTH: 1254  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus equisimilis  
 US-10-172-527-1

Alignment Scores:  
 Pred. No.: 1,976-238 Length: 1254  
 Score: 2145.00 Matches: 417  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 14 Gaps: 0

US-09-469-200D-2 (1-417) x US-10-172-527-1 (1-1254)

QY 1 MetArgThrLeuLysAsnLeuIleThrValValAlaPheSerIlePheTrpValLeuLeu 20  
 Db 1 ATGAGAACATTAATAAAACCTCATAACTGTGTGGCCCTTAGTATATTTTGGGTACTGTGG 60  
 QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40  
 Db 61 ATTTAGTCAATGTTTATCTCTTTGTGCTAAAGGAAGCTTGTCAATTTATGCTTTTGG 120  
 QY 41 LeuIleAlaTyrLeuLeuValLysMetSerLeuSerPhePheTyrIleTyrProPheLysGly 60  
 Db 121 CTGATAGCTTACCTTATAGTCAAAATGCTCTTATCTCTTTTACAAGCCATTAAGGA 180  
 QY 61 ArgAlaGlyGlnTyrLysValAlaIleLeuProSerTyrAsnGluAspAlaGluSer 80  
 Db 181 AGGCTGGCAATATAAGGTTGCAGCCATTATCCCTCTTATAACGAAGATGCTGAGTCA 240  
 QY 81 LeuLeuGluThrLeuLysSerValGlnGlnGlnThrTyrProLeuAlaGluIleTyrVal 100  
 Db 241 TTGCTAGAGACCTTAAAAAGTGTTCAGCAGCAAAACCTATCCCTCAGCAAAATTTATGT 300  
 QY 101 ValAspAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAsp 120  
 Db 301 GTTGAGATGGAAGTGTGATGAGCAGCATTAAGCGCATTAAGCAGTATGCTGCGTGA 360  
 QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140  
 Db 361 ACTGGTGACCTATCAAGCAATGTCTATGTCATCGGTGACAGAAAAATCAAGGAAGCGT 420  
 QY 141 HisAlaGlnAlaTyrAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160  
 Db 421 CATGCACAGCCCTGGGCTTTGAAGATCAGACGCTGATGCTCTTTTACCGCTGACTCA 480  
 QY 161 AspThrTyrIleTyrProAspAlaLeuGluLeuLeuLysThrPheAsnAspProThr 180  
 Db 481 GATACCTATATCTACCTGATGCTTTAGAGGAGTTGTAAAAACCTTTATGACCAACT 540  
 QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200  
 Db 541 GTTTTGTGCGACGGTCACTTAAATGTCAAGAAATAGACAAACCAATCTCTTAACACGC 600  
 QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThr 220  
 Db 601 TTGACAGATTCGCTAATGATATGCTTTGGCTTGAAGAGCTGCCCAATCCGTATCA 660  
 QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValValPro 240  
 Db 661 GGTAATATCTTTGCTCAGCTCGCTTAGCTTTTACAGACGCGAGGTGGTGTCTCT 720  
 QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260  
 Db 721 ACATAGATAGATACATCAACAGACCTCTCTGGGTATTCCTGTAAGTATTTGGTGTATGAC 780  
 QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLys 280  
 Db 781 AGGTGCTTGACCAACTATGCACTGATTTAGGAAGACGTGTTTATCAATCCACTGCTAAA 840  
 QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnAsnArgTrpAsn 300  
 Db 841 TGTATTACAGATGTTCTCTGACAGATGCTACTTACTTGAAGCAGCAAAACCGCTGGAAC 900  
 QY 301 LysSerPhePheArgGluSerIleLeuValLysLysIleMetAsnAsnProPheVal 320  
 Db 901 AGTCTCTTTTAGAGAGTCCATATTCTGTTAGAAATCATGAACATCCCTTTTGTGA 960  
 QY 321 AlaLeuTrpThrIleLeuGluValSerMetPheMetMetLeuValTyrSerValValAsp 340  
 Db 961 GCCCTATGGACCACTCTGAGGTGCTATGTTATGATGCTGTTTATCTGTTGCTGGAT 1020  
 QY 341 PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle 360

QY	101	ValAspAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAsp	120
DB	301	GTTGACGATGAAGTCTGATGACACAGGATTATAGCCATTGAACGACTATGTGCGTGAC	360
QY	121	ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg	140
DB	361	ACTGGTGACCTATCAGCAATGTCTATTGTTCCCGSTCAGAAAAAATCATCAGGAAGCGT	420
QY	141	HisAlaGlnAlaIrrpAlaPheGluArgSerAspAlaAspValPheLeuThrValAspSer	160
DB	421	CATGCACAGGCGTGGCCCTTTCAAGAGATCAGACGGTGATGCTCTTTTGTGACCGCTTGACTCA	480
QY	161	AsnThrThrIleTyrProAsnAlaLeuGlnGluIleuLeuLeuThrProAsnThrProAsnThrDec	180

## 7

Db	481	GATAC	TATATCTAC	CCCTCCAT	CGCTTTTAGAG	AGCTGTTTAAAA	ACCTTTTAA	TAGAC	CCCACT	541		
QY	181	ValPhe	AlaAlaThr	GlyHis	Ileu	AsnValArg	AsnArg	GlnThr	AsnLeu	ThrArg	200	
Db	541	GTTTT	TGTCGG	ACGGGT	CACCTTAA	TGTCAGAA	TATAGACA	AAACCA	TATCTTTAA	CACGC	600	
QY	201	LeuThr	AspIle	ArgTyr	AspAsn	AlaPhe	GlyVal	GluArg	AlaAla	GlnSer	ValThr	220
Db	601	TTGAC	AGATAT	TCGCTAT	CA	TAAATG	CTTTTGG	CGGTTGA	ACGAGCT	CCCATCC	GTGTACA	660
QY	221	GlyAsn	IleLeu	ValCys	SerGly	ProLeu	SerValTyr	ArgArg	GluVal	ValVal	Pro	240
Db	661	GGTA	ATATTC	TCGTTG	CTCAG	CGCCGCTTAG	GGTTTAC	AGACG	CGAGGT	GTTGTTCT	720	
QY	241	AsnIle	AspArg	TyrGly	IleAsn	GlnThr	PheLeu	GlyIle	ProVal	SerIle	CysAsp	260
Db	721	ACAT	GATAGAT	ATACAT	CAAC	CAACCTTCC	TGGGTAT	TCTCTAA	TGATCG	GTGTAG	780	
QY	261	ArgCys	LeuThr	AsnTyr	AlaThr	AspLeu	GlyLys	ThrValTyr	GlnSer	ThrAla	Lys	280
Db	781	AGGT	GCTTG	ACCACTAT	CCAACTG	ATTAG	NAAGACTG	TTTATCAT	CCAC	TCGTAA	840	
QY	281	CysIle	ThrAsp	ValPro	AspLys	MetSer	ThrTyr	LeuLys	GlnGln	AsnArg	TrpPAsn	300
Db	841	TGTAT	TACAGAT	GTCTCT	CACA	AAGATG	CTCTACT	TACTTGA	AGCAGCA	AAACCG	CTGGAC	900
QY	301	LysSer	PhePhe	ArgGlu	SerIle	IleSer	ValLys	LysLys	IleMet	AsnAsn	ProPhe	320

**Db**

QY	321	AlaLeuTrpThrIleuGluValSerMetPheMetMetLeuValIleSerValValAsp	340
Db	961	GCCCATGGACCATCTTCAGGTGCTCTATTTATGATGCTGTTCTTTATCTGTGGAT	1021
QY	341	PhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle	360

D5

```

QY      361   IlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPhe 380
          |||||
Db      1081 ATCTCATGTGCTCTTTGGCGAAATACACTATATGCTTGAAGCACCOCCTGCCTTC 1147
          |||||
QY      381   LeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyr 400
          |||||
Db      1141 TTGTTATCTCCGTTTTATGGSGTACTGCATTGTTGTTGCTACAGCCCTTCAAATGTAT 1200
          |||||
QY      401   SerLeuPheThrIleArgAsnAlaaspTrpGlyThrArgLysLysLeuLeu 417
          |||||
Db      1201 TCTCTTTTACTATTAGAAAAGCTGACTGGGAGCACGTAATAAATATTATTA 1251

RESULT 6
US-10-326-185-108
; Sequence 108, Application US/10326185
; Publication No. US20030175902A1
; GENERAL INFORMATION:
; APPLICANT: Sioma, Alan
; APPLICANT: Behr, Regine
; APPLICANT: Widner, William

```

; APPLICANT: Sternberg, David  
 ; APPLICANT: Brown, Stephen  
 ; TITLE OF INVENTION: Methods for Producing Hyaluronan In a Recombinant Host Cell  
 ; FILE REFERENCE: 10241.200-US  
 ; CURRENT APPLICATION NUMBER: US/10/326,185  
 ; CURRENT FILING DATE: 2002-12-20  
 ; PRIOR APPLICATION NUMBER: US 60/342,644  
 ; PRIOR FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 108  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 108  
 ; LENGTH: 5158  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus equisimilis  
 US-10-326-185-108

Alignment Scores:  
 Pred. No.: 8,48e-218 Length: 5158  
 Score: 1976.00 Matches: 383  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 92.12% Indels: 0  
 Gaps: 12

US-09-469-200D-2 (1-417) x US-10-326-185-108 (1-5158)

QY 35 SerileTyrGlyPheLeuLeuLeuAlaTyrLeuLeuValLysMetSerLeuSerPhePhe 54  
 Db 1 TCAATTTATGGCTTTTGTCTGATAGCTTACCTATTAGTCAAAATGTCCTTATCCCTTTT 60  
 QY 55 TyrLysProPheLysGlyArgAlaGlyGlnTyrLysValAlaLeuLeuLeuLeuLeu 74  
 Db 61 TACAAGCCATTTAAGGAAGGCTGGCAATATAGGTTGCAGCCATTTATCCCTCTAT 120  
 QY 75 AsnGluAspAlaGluSerLeuLeuGluThrLeuLysSerValGlnGlnGlnThrPro 94  
 Db 121 AACCAAGATCTGAGTCAATCTGAGACCTTAAAGTGTTCAGCAGCAACCTATCCC 180  
 QY 95 LeuAlaGluLeuTyrValValAspGlySerAlaAspGluThrGlyLeuLysArgTle 114  
 Db 181 CTAGCAGAAATTTATGTTGACGATGGAGTGTGATGACAGATTAAGCGCAT 240  
 QY 115 GluAspTyrValArgAspThrGlyAspLeuSerSerAsnValLeuHisArgSerGlu 134  
 Db 241 CAAGACATATGCGTGACACTGGTGACCTATCAAGCAATGTCATTCATCGGTGAG 300  
 QY 135 LysAsnGlnGlyArgHisAlaGlnAlaTyrAlaPheGluArgSerAspAlaAspVal 154  
 Db 301 AAAATCAAGAAAGCGTCATGACAGGCTGGGCTTTGAAGATCAGACGCTGATGTC 360  
 QY 155 PheLeuThrValAspSerAspThrTyrTyrProAspAlaLeuGluLeuLeuLys 174  
 Db 361 TTTTGTGACCGTGACTCAGATACATATATCTACCTGATGCTTTAGAGGAGTTGTTAAA 420  
 QY 175 ThrPheAsnAspProThrValPheAlaAlaThrGlyHisLeuAsnValArgAsnGln 194  
 Db 421 ACCCTTATGACCAACTGTTTGTGCGACGCGGTACCTTAAATGCTCAGAAATAGACA 480  
 QY 195 ThrAsnLeuLeuThrArgLeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArg 214  
 Db 481 ACCAATCTCTACACGCTTGCAGATATTCGCTATGATATGCTTTTGGCGTTGACGA 540  
 QY 215 AlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGlyProLeuSerValTyrArg 234  
 Db 541 GCTGCCAATCCGTTACAGGTAATATCTTTGCTCAGGTCGGCTTAGCGTTTACAGA 600  
 QY 235 ArgGluValValProAsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIlePro 254  
 Db 601 CCGGAGGTGTTTCTTCAACATAGATAGATACATCAACACGACCTTCTCGGTATTCCT 660  
 QY 255 ValSerIleGlyAspAspArgCysLeuThrAsnTyrAlaThrAspLeuGlyIleThrVal 274  
 Db 661 GTAAGTATGGTGATGACAGGTGCTTGACCACTATGCACTGATTTAGGAAGACTGTT 720

QY 275 TyrGlnSerThrAlaLysCysIleThrAspValProAspLysMetSerThrTyrLeuLys 294  
 Db 721 TATCAATCCACTGCTAAATGATATTACAGATGTTCTCTGACAAGATGCTTACTTACTTGAAG 780  
 QY 295 GlnGlnAsnArgTyrAsnLysSerPheArgGluSerIleIleSerValLysIle 314  
 Db 781 CAGCAAAACCGCTGGAACAAGTCTTTTAGAGAGTCCATTAATTCGTGAAGAAATC 840  
 QY 315 MetAsnAsnProPheValAlaLeuLeuThrIleLeuGluValSerMetPheMetMetLeu 334  
 Db 841 ATGACAAATCTTTTGTAGCCCTATGGACCATCTAGAGTGTCTATGTTTATGATGCTT 900  
 QY 335 ValTyrSerValValAspPheValGlyAsnValArgGluPheAspTyrLeuArgVal 354  
 Db 901 GTTTATTCGTGGTGGATTTCTTTTAGCAATGTCAGAGAAATTTGATGGCTCAGGGTT 960  
 QY 355 LeuAlaPheLeuValIleIlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeu 374  
 Db 961 TTAGCCCTTTCTGGTGATTAICTTCAITGTTGCCCTGTGCGGACATTCATTACATGCTT 1020  
 QY 375 LysHisProLeuSerPheLeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeu 394  
 Db 1021 AAGCACCCGCTGCTCTCTTGTATCTCCGTTTATGGGTGCTGCATTTGTTTGTCTTA 1080  
 QY 395 GlnProLeuLysLeuTyrSerLeuPheThrIleArgAsnAlaAspTyrGlyThrArgLys 414  
 Db 1081 CAGCCCTTGAATTAATATCTCTTTTACTATTAGAAATGCTGACTGGGAACACGTAAA 1140  
 QY 415 LysLeuLeu 417  
 Db 1141 AAATTATTA 1149

# RESULT 7

US-10-326-185-102  
 ; Sequence 102, Application US/10326185  
 ; Publication No. US20030175902A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sloma, Alan  
 ; APPLICANT: Behr, Regine  
 ; APPLICANT: Widner, William  
 ; APPLICANT: Tang, Maria  
 ; APPLICANT: Sternberg, David  
 ; APPLICANT: Brown, Stephen  
 ; TITLE OF INVENTION: Methods for Producing Hyaluronan In a Recombinant Host Cell  
 ; FILE REFERENCE: 10241.200-US  
 ; CURRENT APPLICATION NUMBER: US/10/326,185  
 ; CURRENT FILING DATE: 2002-12-20  
 ; PRIOR APPLICATION NUMBER: US 60/342,644  
 ; PRIOR FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 108  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 102  
 ; LENGTH: 1251  
 ; TYPE: DNA  
 ; ORGANISM: Streptococcus uberis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1251)  
 ; OTHER INFORMATION:  
 US-10-326-185-102

Alignment Scores:  
 Pred. No.: 2,67e-186 Length: 1251  
 Score: 1695.50 Matches: 312  
 Percent Similarity: 87.47% Conservative: 51  
 Best Local Similarity: 75.18% Mismatches: 51  
 Query Match: 79.04% Indels: 1  
 Gaps: 12

US-09-469-200D-2 (1-417) x US-10-326-185-102 (1-1251)

QY 1 MetArgThrLeuLysAsnLeuIleThrValValAlaPheSerIlePheTrpValLeu 20

Db 1 ATGGAACAACTAAAAATCTCATTTATGACATTTATTTCCCTGCGCTCAATAT 60  
QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40  
Db 61 ATTGGGCTTAATGTTTGTATTTGGAACCTAAAGGAAGTCTAACAGGTGTATGGATATT 120  
QY 41 LeuIleAlaTyrLeuValLysMetSerLeuSerPhePheTyrLysProPheLysGly 60  
Db 121 CTATTAACTATTGTCGATAAAATGGGATTATCTTTTATTCGTCCTATAAAGA 180  
QY 61 ArgAlaGlyGlnTyrLysValAlaIleProSerTyrAsnGluAspAlaGluSer 80  
Db 181 AGTGTAGTCAATATAGTACGCTATTATCCATCTATATAGGATGGTCTCGGT 240  
QY 81 LeuLeuGluThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVal 100  
Db 241 TTACTAGAACTCTAAAGAGTGTTCAAAACAAACATATCCAAATGTCAGAAATTTTCGTA 300  
QY 101 ValAspAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAsp 120  
Db 301 ATTGACGATGGTCAAGTATATAAACAGGTATATAATGTCGAGACTATGTAAGTTA 360  
QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140  
Db 361 AATGGC---TTTGAGACCAAGTATATCGTTCATCAGATCCCTGAAATGTTGTAAGA 417  
QY 141 HisAlaGlnAlaTyrPheGluArgSerAspAlaAspValPheLeuThrValAspSer 160  
Db 418 CAGTCTCAGGCTGGGATTTGAAAGGCTGATGCTGATGTTTCTTAAACAGTGGATCA 477  
QY 161 AspThrTyrIleTyrProAspAlaLeuGluLeuLeuLysThrPheAsnAspProThr 180  
Db 478 GATACCTACATCTATCTGATGCTCTTCAGAAATTTATAAGACATTTATGATCCAGAG 537  
QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200  
Db 538 GTCTAGCTGCACTGCTGCTATTAATGCAAGAAATAGACAACTAATCTCTTAACATA 597  
QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaGlnSerValThr 220  
Db 598 CTGCTGATATCTGCTATACATATGCTATGCTGAGACGCTGCTCAGTCTGTTAGC 657  
QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgArgGluValValPro 240  
Db 658 GGAATATTTTGGTTGTCGACCTTTAAGTATTTATAGACGTCCTCGGTATCCA 717  
QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260  
Db 718 AATCTGTAACGCTATACCTCACAACATTTCTTGGTGTCCCTGTAAGCATGGGATGAC 777  
QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGlyLysThrValTyrGlnSerThrAlaLys 280  
Db 778 CGTGTGTGCACAAATATGCACTGATTTGGGAAAAACGGTTATCAGTCACTGCAAGA 837  
QY 281 CysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTyrAsn 300  
Db 838 TGTGATCTGACGTTCCAGATAAGTTTAAAGGTTTTCATCAACAAACAAATCTGTGAT 897  
QY 301 LysSerPhePheArgGluSerIleIleSerValLysLysIleMetAsnAsnProPheVal 320  
Db 898 AAGTCATTTTATAGGAGTCTATATCTCTTAAAGATTATTAGCCACCAACCAAGTGT 957  
QY 321 AlaLeuIleThrIleLeuGluValSerMetPheMetLeuValTyrSerValAsp 340  
Db 958 CGCTGTGAGCTATTACAGAAAGTTTCCATGTTTCATCATGCTAGTTTATCTATCTTAC 1017  
QY 341 PhePheValGlyAsnValArgGluPheAspTyrLeuArgValLeuAlaPheLeuValIle 360  
Db 1018 TTATGTAGGAGGCTCAAGAATTTAATCTCATAAACTGGTGTCTTTTAGTTAT 1077  
QY 361 IlePheIleValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSerPhe 380

Db 1078 ATTTTCATAGTACGCTTTTGTAGAAATGTTTACATGCTTAAGCATCCATTTGCTTTT 1137  
QY 381 LeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyr 400  
Db 1138 TTATGTACCGCTTTTATGATTGATACATCTATTGCTTTTGCACCTCTTAAGATATAT 1197  
QY 401 SerLeuPheThrIleArgAsnAlaAspTyrGlyThrArgLysLys 415  
Db 1198 TCGTATTACTATAGAATGCTACATGGGAACCTGTAAGAAG 1242

RESULT 8  
US-10-172-527-11  
; Sequence 11, Application US/10172527  
; Publication No. US20030092118A1  
; GENERAL INFORMATION:  
; APPLICANT: Weigel, Paul H  
; APPLICANT: Kumari, Kshama  
; APPLICANT: DeAngelis, Paul  
; TITLE OF INVENTION: HYALURONAN SYNTHASE GENES AND EXPRESSION THEREOF IN BACILLUS S  
; FILE REFERENCE: 3554 048  
; CURRENT APPLICATION NUMBER: US/10/172, 527  
; CURRENT FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: 60/297,788  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: 60/297,744  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: 09/469, 200  
; PRIOR FILING DATE: 1999-12-21  
; PRIOR APPLICATION NUMBER: 09/178,851  
; PRIOR FILING DATE: 1998-10-26  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11  
; LENGTH: 3466  
; TYPE: DNA  
; ORGANISM: Streptococcus uberis  
US-10-172-527-11

Alignment Scores:  
Pred. No.: 1,48e-185 Length: 3466  
Score: 1695.50 Matches: 312  
Percent Similarity: 87.47% Conservative: 51  
Best Local Similarity: 75.18% Mismatches: 51  
Query Match: 79.04% Indels: 1  
DB: 14 Gaps: 1

US-09-469-200D-2 (1-417) x US-10-172-527-11 (1-3466)

QY 1 MetArgThrLeuLysAsnLeuIleThrValValAlaPheSerIlePheThrPValLeuLeu 20  
Db 23 ATGGAACAACTAAAAATCTCATTTATGACATTTATTTCCCTGCGCTCAATAT 82  
QY 21 IleTyrValAsnValTyrLeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeu 40  
Db 83 ATTGGGCTTAATGTTTGTATTTGGAACCTAAAGGAAGTCTAACAGGTGTATGGATATT 142  
QY 41 LeuIleAlaTyrLeuValLysMetSerLeuSerPhePheTyrLysProPheLysGly 60  
Db 143 CTATTAACTATTGTCGATAAAATGGGATTATCTTTTATTCGTCCTATAAAGA 202  
QY 61 ArgAlaGlyGlnTyrLysValAlaIleProSerTyrAsnGluAspAlaGluSer 80  
Db 203 AGTGTAGTCAATATAGTACGCTATTATCCATCTTATAATGAGATGGTGTGCGGT 262  
QY 81 LeuLeuGluThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVal 100  
Db 263 TTACTAGAACTCTAAAGAGTGTTCAAAACAAACATATCCAAATGTCAGAAATTTTCGTA 322  
QY 101 ValAspAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAsp 120  
Db 323 ATTGACGATGGTCAAGTATATAAACAGGTATAAATGTCGAGACTATGTAAGTTA 382  
QY 121 ThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGlnGlyLysArg 140



US-09-469-200D-2 (1-417) x US-10-172-527-13 (1-1440)

[illegible]

US-10-117-795-1

Application US/10117795  
O. US2003Q10453A1  
NATION:  
EIGEL, PAUL E.  
DEANGELIS, PAUL L.  
JOHN  
PAPACONSTANTINO,  
VENTION: HYALURONATE SYNTHASE GENES AND USES THEREOF  
CE: 35541.073

US-09-469-200D-2 (1-417) x US-10-124-222-1 (1-1512)

[illegible]

1402 ATTTAAATACGGAATGGGAACACGTAAAAAGTC 1437  
 Db  
 CATTTTAAGGAAATCCACATGAC 381

US-10-117-795-1  
; Sequence 1, Application US/10117795  
; Publication No. US2003010453A1

luiletytValvalaspaspGly 104  
 |||||:|||||||  
 ; APPLICANT: WEIGEL, PAUL F.  
 ; APPLICANT: DEANGELIS, PAUL L.  
 |||||:|||||||  
 ; APPLICANT: PAPAConstantinou, John  
 AAAATATATGTTGATGATGG 501

yrValArgAspThrGlyAspLeu 124  
; TITLE OF INVENTION: HYALURONATE SYNTHASE GENES AND USES THEREOF  
; FILE REFERENCE: 35541.073

Db 802 CGTTACGATAATGCCCTTTGGGGTGGAGCGTCTCAATCATTAACAGGTAATATTTA 861

Best Local Similarity: 51.23% Mismatches: 98  
 Query Match: 45.57% Indels: 17  
 DB: 14 Gaps: 2

US-09-469-200D-2 (1-417) x US-10-172-527-17 (1-1200)

QY 66 LysValAlaAlaIleProSerTyrAsnGluAspAlaGluSerLeuLeuGluThrLeu 85  
 DB 17 AAAGTGGCAGTAGTGGTCCCTCATATATGAAGTGAAGTGGTATTTAATCAAT 76  
 QY 86 LysSerValGlnGlnThrTyrProLeuAlaGluIleTyrValValAspAspGlySer 105  
 DB 77 AATAGCGTTTATAGTCAAGATTATCAATATCAATATTTCTTTGTTGATGATGTAGT 136  
 QY 106 AlaAsp-----GluThrGlyIleLysArgIleGluAspTyrValArgAspThr 121  
 DB 137 AAGATAAATTCGGCTTATGAAGTAGCACTTAAATGAGGAGGAATCTTGAAGACCAA 196  
 QY 122 GlyAspLeuSerSer-----AsnVal 128  
 DB 197 CGAAGAAATTCCTGCTACAACTAAGAAATTTCTGTAATATATAGTATTCCTCACTTA 256  
 QY 129 IleValHisArgSerGluLysAsnGlnGlyLysArgHisAlaGlnAlaIlePalaPheLeu 148  
 DB 257 ATTCGTATCGTTTATAGTAAATTCGCGAAGAACATGCTCAATATATGGCTTTAA 316  
 QY 149 ArgSerAspAlaAspValPheLeuThrValAspSerAspThrTyrIleTyrProAspAla 168  
 DB 317 CGGACACACAGATGCTATTTGACCATGATGATGATGATGATGATGATGATGATGAT 376  
 QY 169 LeuGluGluLeuLeuLysThrPheAsnAspProThrValPheAlaAlaThrGlyHisLeu 188  
 DB 377 GTTAGAGAGTATTGAACCCCTTAAGTGAAGAAAGTAAATGGCCACAACTGTCACGTG 436  
 QY 189 AsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIleArgTyrAspAsn 208  
 DB 437 ACATTCGTATAGAAATGATTAATTAACAACTAATGATGATGATGATGATGATGATGAT 496  
 QY 209 AlaPheGlyValGluArgAlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGly 228  
 DB 497 GCGTCCGTGTGGAGCGTGCAGCACAGTCCGTAACAGAAATGTTCTGTTGTAGTGGG 556  
 QY 229 ProLeuSerValThrArgGluValValProAsnIleAspArgTyrIleAsnGln 248  
 DB 557 CGGTAAAGTGTATTCGTAGAGAAATTAATCACTGCTGATGATGATGATGATGATGATGAT 616  
 QY 249 ThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThrAsnTyrAlaThr 268  
 DB 617 ATGTTCTTGTGAGAGGTGAGTGGAGATGATGATGATGATGATGATGATGATGATGAT 676  
 QY 269 AspLeuGlyLysThrValThrGlnSerThrAlaLysCysIleThrAspValProAspLys 288  
 DB 677 TTGAAGGGAACACAGTTATCAATCCACGCTGCTGATGATGATGATGATGATGATGATGAT 736  
 QY 289 MetSerThrTyrLeuLysGlnGlnAsnArgTyrAsnLysSerPhePheArgGluSerIle 308  
 DB 737 TTAACAACATTTCTTAACAGCAACTACGTGGAACAAAGTCAATTTTATGAGAAAGTTTA 796  
 QY 309 IleSerValLysLysIleMetAsnPropheValAlaLeuThrThrIleLeuGluVal 328  
 DB 797 ATTTCACCTGGCATTTGGTATGAAACCAATGTTCTGTTGGACATTTTCGAAATA 856  
 QY 329 SerMetPheMetMetLeuValTyrSerValValAspPhePheValGlyAsnValArgGlu 348  
 DB 857 TCGTTATGGATTTATTTGGGCTTTCCTACTTCTAAGTATTTATTCACAGGCAAGTCA 916  
 QY 349 PheAspTrpLeuArgValLeuAlaPheLeuValIleIlePheIleValAlaLeuCysArg 368  
 DB 917 GTAGGGTTAATTTGGCTGTTTATTTATTTGGGTTATATTTATTTATTTATTTATTTAT 976  
 QY 369 AsnIleHisTyrMetLeuLysHisProLeuSerPhePheLeuLeuSerProPheTyrGlyVal 388  
 DB 977 AATGATTTATCTATTAACATCCCTTACTTCTTACTTCTTACTTCTTACTTCTTACTTACT 1036

QY 389 LeuHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThrIleArgAsnAla 408  
 DB 1037 CTCATGATTAGCACTATTACCTATAGCTTTTATGCTTACTTAACATAATTAATCTAAT 1096  
 QY 409 AspTyrGlyThrArg 413  
 DB 1097 GGTGGGGAACAGT 1111  
 RESULT 14  
 US-09-939-964-1/c  
 ; Sequence 1, Application US/09939964  
 ; Publication No. US20030054522A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosenbhal, Andre  
 ; APPLICANT: Freiberg, Christoph  
 ; APPLICANT: Perret, Xavier Philippe  
 ; APPLICANT: Broughton, William John  
 ; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic  
 ; TITLE OF INVENTION: Plasmid  
 ; FILE REFERENCE: CARP0068  
 ; CURRENT APPLICATION NUMBER: US/09/939,964  
 ; CURRENT FILING DATE: 2001-08-27  
 ; PRIOR APPLICATION NUMBER: 09/214,808  
 ; PRIOR FILING DATE: 1999-06-22  
 ; NUMBER OF SEQ ID NOS: 1  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 536165  
 ; TYPE: DNA  
 ; ORGANISM: Rhizobium  
 US-09-939-964-1  
 Alignment Scores:  
 Pred. No.: 6,37e-39 Length: 536165  
 Score: 463.00 Matches: 131  
 Percent Similarity: 48.75% Conservative: 64  
 Best Local Similarity: 32.75% Mismatches: 185  
 Query Match: 21.59% Indels: 20  
 DB: 11 Gaps: 10  
 US-09-469-200D-2 (1-417) x US-09-939-964-1 (1-536165)  
 QY 27 LeuPheGlyAlaLysGly-----SerLeuSerIleTyrGlyPheLeuLeuIleAlaTyr 44  
 DB 157330 CTGCTTGGCAGCAGCGCGGTAGCCATCTCTGTATGACAGCAGCTCTCGAGCGCTTAC 157271  
 QY 45 -----LeuLeuValLysMetSerLeuSerPhePheTyrLysProPhe 58  
 DB 157270 AAAGCGATGCAAGCTATATACGTTTGGCAACACACAGCTGCTGCAAGCGCGGTG 157211  
 QY 59 LysGlyArgAlaGlyGlnTyrLysValAlaAlaIleIleProSerTyrAsnGluAspAla 78  
 DB 157210 ACCGGTCCGTGGTGCACCGCGAGCTGATGATGATGATGATGATGATGATGATGATG 157151  
 QY 79 GluSerLeuLeuGluThrLeuLysSerValGlnGlnGlnThrTyrPro---LeuAlaGlu 97  
 DB 157150 CGCGCGCTCTCGCGGTGCTGCTTCCATTCGAAAGCAGGAGCTACGCTGAGAGTGGCG 157091  
 QY 98 IleTyrValValAspAspGlySerAlaAspGluThrGlyIleLysArgIleGluAspThr 117  
 DB 157090 GTCTAGCTGGTGTGACAGCGGTCTTGGCAATCGCAACGCCATCATCTGTACAGCATCAT 157031  
 QY 118 ValArgAspThrGlyAspLeuSerSerAsnValIleValHisArgSerGluLysAsnGln 137  
 DB 157030 TAT-----GGTGGCAGCGAGGTTCGCTTTATCTCTG-----ATGCCAAAGAACGTC 156983  
 QY 138 GlyLysArgHisAlaGlnAlaIlePalaPheGluArgSerAspAlaAspValPheLeuThr 157  
 DB 156982 GGAAGCGCAAGCGAGATGTCGCGATACGGGAATCATCGGAGATTTGGTGTCTCAAC 156923  
 QY 158 ValAspSerAspThrTyrIleTyrProAspAlaLeuGluLeuLeuLysThrPheAsn 177





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 2, 2003, 12:39:19 ; Search time 2815 Seconds  
(without alignments)  
3600.343 Million cell updates/sec

Title: US-09-469-200D-2  
Perfect score: 2145  
Sequence: 1 MTKLNLTIVAFSIFWLL.....KLYSLTIRNADGTRKLL 417

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q/cgn2.1/USPTO.spool/US09469200/runat\_02102003\_090348\_22788/app\_query.fasta.1.583  
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -IOOCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MAFEX=blotsum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09469200.@CGN.1.2810 @runat\_02102003\_090348\_22788 -NCPU=6  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estim:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gssi:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	440	20.5	4151	11	AK028582	Mus muscu
2	439	20.5	2891	11	AK079729	Mus muscu
3	320	14.9	809	14	CA324064	UI-M-FY0-
4	313.5	14.6	980	29	BZ606767	BZ606767 WHADEIOTF
5	297.5	13.9	711	13	B0278152	B0278152 603865094
6	291	13.6	1084	12	BM544718	BM544718 AGENCOURT
7	286	13.3	915	13	BM371636	BM371636 BX371636
8	284	13.2	639	9	AL775141	AL775141 AL775141
9	280	13.1	551	10	BG707429	BG707429 602672848
10	280	13.1	551	10	BG707608	BG707608 602670872
11	273.5	12.8	656	12	BJ041963	BJ041963 BJ041963
12	273	12.7	536	12	BJ062498	BJ062498 BJ062498
13	267.5	12.5	954	13	B0910636	B0910636 AGENCOURT
14	263	12.3	866	14	CA988122	CA988122 AGENCOURT
15	262.5	12.2	876	13	B0907404	B0907404 AGENCOURT
16	261.5	12.2	782	12	B1753116	B1753116 603025961
17	260	12.1	857	13	B0906272	B0906272 AGENCOURT
18	259.5	12.1	948	13	BX327795	BX327795 BX327795
19	245	11.4	623	9	AL966439	AL966439 AL966439
20	242.5	11.3	672	12	BJ073147	BJ073147 BJ073147
21	236	11.0	546	10	BF193947	BF193947 245690 MA
22	224.5	10.5	699	9	AL647479	AL647479 AL647479
23	222	10.3	643	28	AZ282684	AZ282684 RC21-23-1
24	215.5	10.0	883	14	CA793387	CA793387 AGENCOURT
25	199.5	9.3	704	12	BH066629	BH066629 BJ066629
26	189	8.8	481	2	HSM089391	HSM089391 Homo sapl
27	187.5	8.7	517	10	BG348186	BG348186 dg39503.Y
28	186.5	8.7	542	10	BG364756	BG364756 dg35f08.Y
29	185	8.6	1069	29	CNS0682W	AL412626 T7 end of
30	184.5	8.5	991	13	BQ725516	BQ725516 AGENCOURT
31	182	8.5	923	13	B0909569	B0909569 AGENCOURT
32	181.5	8.5	885	14	CA987219	CA987219 AGENCOURT
33	178	8.3	1135	29	CNS06522	AL413892 T3 end of
34	177.5	8.3	376	9	AU244838	AU244838 AU244838
35	176	8.2	895	13	B0911059	B0911059 AGENCOURT
36	173.5	8.1	548	10	BG347935	BG347935 de71e03.Y
37	170	7.9	862	29	CNS03W09	AL264186 Tetraodon
38	169	7.9	859	14	CA986884	CA986884 AGENCOURT
39	164	7.6	485	13	BQ828135	BQ828135 LL61n1142
40	162	7.6	482	12	BG892747	BG892747 de54f11.Y
41	155.5	7.2	869	14	CD052715	CD052715 LITHZF000
42	155.5	7.2	941	29	CNS069FT	AL413199 T7 end of
43	154.5	7.2	533	9	AL924720	AL924720 AL924720
44	154	7.2	750	28	BH367878	BH367878 AG-ND-124
45	153.5	7.2	831	14	CA986525	CA986525 AGENCOURT

# ALIGNMENTS

RESULT 1  
AK028582  
LOCUS  
DEFINITION  
Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:4732404.04 product:similar to D642111 [Brachydanio rerio], full insert sequence.  
ACCESSION  
AK028582  
VERSION  
AK028582.1 GI:26324525  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

## REFERENCE

- 1 Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 11049636
- 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159
- 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakauchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system-384 format sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 PUBMED 11076861
- 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, K., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleschmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staab, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, I., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weltz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409 (6821), 685-690 (2001)  
 MEDLINE 21085660  
 PUBMED 11217851
- 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 MEDLINE 12141511
- 6 (bases 1 to 4151)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hayashizaki, Y., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akshira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://fantom.gsc.riken.go.jp/.

## COMMENT

## FEATURES

## Location/Qualifiers

1..4151

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM\_DB:4732404L04"

/db\_xref="taxon:10090"

/clone="4732404L04"

/tissue\_type="skin"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="10 days neonate"

118..1782

/note="unnamed protein product; putative

similar to DG42III [Brachydanio rerio] (SPTR|09DG40,

evidence: FASTA, 75.1%id, 99.8%length, match=1659)\*

/codon\_start=1

/protein\_id="BAC26017.1"

/db\_xref="GI:26324526"

/translation="MPVLTALRVVGTSLFVVLGGILAAVYGVQFIHTEKHVLS

FGLVAILGHLIIQSLFAFLHRRMERACRPLKHCSPRSPRVALCICIAVEDPEYL

RCLRSORAFPNLKVWVVDNEDYMLDIEHVGTCQAGFFWRSNHEAG

EEVEASLQSGMERVAVYVNSTFSCIMKNGKREVMITAFALGNSVDYIQVCSGD

TVDPQCTISGLMVRLEEDPOVGGDVLNKKYDSWISFLSVRYMAFNVACROS

YFGVQCTISGLMVRLEEDPOVGGDVLNKKYDSWISFLSVRYMAFNVACROS

TARSKCTETPRYLNQTLRMSKYPRELYNSLWFKHLMWTYSVVTGFFPF

FLIATVQLFGRYINLLFLVQLVQIKATYACFLGNAMFLMFLSYLLVMS

LPKATIAIATINKSGMTSGRKTIVNFGILPVSIVWVLLGLLAYTAYQDLFSE

TLAPLVSGAILYGCYVWALLMLYLIIARRGKKPEQYSIAFV"

4131..4136

/note="putative"

polyA\_signal

polyA\_site

BASE COUNT 983 a 1013 c 1055 g 1100 t

ORIGIN

Alignment Scores:

Pred. No.: 1.39e-36 Length: 4151

Score: 440.00 Matches: 134

Percent Similarity: 44.44% Conservative: 86

Best Local Similarity: 27.07% Mismatches: 171

Query Match: 20.51% Indels: 104

DB: 11 Gaps: 18

US-09-469-200D-2 (1-417) x AK028582 (1-4151)

QY 4 LeuLysAsnLeuLeuThrValValAlaPheSerIlePhe-----Ttp 17

Db 130 CTGACTACAGCCCTGGCTGGTGGGCGACAGCTCTGTTGCCCTGTAGTCTGGGAGGC 189

QY 18 ValLeuLeuLeuTyrValAsnValTyrLeuPhe-----GlyAlaLysGlySerLeuSer 35

Db 190 ATCTGGCGGCTATGTGACAGGCTACCATGTTTATCCACACAGAAAGCACTACCTGTCC 249

QY 36 -----IleTyrGlyPheLeuLeuLeuAlaTyrLeuValLysMetSerLeuSerPhe 53

Db 250 TTGGGCTCTACGCTGGCCATCTCGGCTGTACATCTGCTCATCCAGAGCCTGTTGCTTC 309

QY 54 PheTyrLysProPheLysGlyArgAlaGlyGln----- 64

Db 310 CTGGACACCCCTCCGAAATGCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 369

QY 65 ---TyrLysValAlaAlaIleIleProSerTyrAsnGluAspAlaGluSerLeuLeuGlu 83  
 Db 370 CGCGCTTCAGTGGCCTCGCTACCAAGAGGACCCGCAAGTACCTGGCGAAG 429  
 QY 84 ThrLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrValValAspAsp 103  
 Db 430 TGCCTTCGCTCAGCTCAGCGCATTCCTTCCAAACCTCAAGGTGGTCATGGTGGAT 489  
 QY 104 GlySerAlaAspGluThrGlyIleLysArgIleGluAspTyrValArgAspThr 121  
 Db 490 GGCATCGCCAGGAA-----GATACCTACATGTGTGGACATCTTCCAT 531  
 QY 122 -----GlyAspLeuSerSerAsnValIleValHisArgSer-----Glu 134  
 Db 532 GAGGTGCTGGTGGCAGCTCAGCAGCTCGCTTCTTGTGGCTGACCAATTCATGAG 591  
 QY 135 LysAsnGlnGlyLysArgHisAla-----142  
 Db 592 GCGGTGAAGGAGACAGAGCCGCTGCAGGAAGCATGGAGCGTGGCGAGCTGTG 651  
 QY 143 -----GlnAlaTrpAlaPheGluArg-----149  
 Db 652 GTGTGGCCGACACCTTCATGATCATCAGAAAGTGGGGGCGCAGCGTGGATCATG 711  
 QY 150 -----SerAspAlaAspValPheLeuThrValAspSerAsp 161  
 Db 712 TACACTGCCTCAAGCCCTTGGCACTCAGTGGCTACATCCAGGTGTGCTACTCTGAC 771  
 QY 162 ThrTyrIleTyrProAspAlaLeuGluLeuLysThrPhe---AsnAspProThr 180  
 Db 772 ACTGTGTGGCAGCCCTGCACCATTCAGATTCGATGCTTGGAGAAAGATCCCAA 831  
 QY 181 ValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArg 200  
 Db 832 GTAGAGGTGTGGAGAGATGCTCAATCTCCACAGATGATGATCATGGATCTCCCTC 891  
 QY 201 LeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThr 220  
 Db 892 CTGACGAGTGTGAGTGTGATGGCTTTCACGTGGAGCGGCTGCCAGTCTACTTT 951  
 QY 221 GlyAsnIleLeuValCysSerGlyProLeuSerValTyrArgGluValValValPro 240  
 Db 952 GCGCTGTGCAATGATTAATAGTGGCCCTTGGCATGTACCCGACACGCTCTCCAGAG 1011  
 QY 241 AsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAsp 260  
 Db 1012 TTCTGGAGGATGGTACCATCAGAGTCTCTAGGACGACAGTGCACCTTTGGGATGAT 1071  
 QY 261 ArgCysLeuThrAsnTyrAlaThrAspLeuGly---LysThrValTyrGlnSerThrAla 279  
 Db 1072 CGGCACCTTACCACCGAGTCTGTGAGTCTTGGCTACCGAGTAAAGTATACAGCGCTCT 1131  
 QY 280 LysCysIleThrAspValProAspLysMetSerThrTyrLeuLysGlnAsnArgTrp 299  
 Db 1132 AGTGCTTCACAGAGACCCCACTAGTACCTTCGATGGCTCAATCAGCAACCCGCTGG 1191  
 QY 300 AsnLysSerPhePheArgGluSerIle-----308  
 Db 1192 ACCAAGTCTTACTTCGGGAATGGCTCTACAAATCTCTGTGTTCATAGCACCACTC 1251  
 QY 309 ---IleSerValLysLysIleMetAsnAsp-----ProPheValAlaLeuTrpThrIle 325  
 Db 1252 TGGATGACCTATGAATCAGTGGTGCACAGGTTCTTCCCATTCCTTCATTCACATGCT 1311  
 QY 326 LeuGluValSerMetPheMetMetLeuValTyrSerValValAspPhePheValGlyAsn 345  
 Db 1312 ATACAACTT-----TTTACCGTGGCGCG 1335  
 QY 346 ValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIleIlePheIleValAla 365  
 Db 1336 ATC-----TGG---NACATTCTCTCTCTCAACAGTGCAGCTGGTGGCG 1380

QY 366 LeuCysArgAsnIleHis---TyrMetLeuLysHisProLeuSerPheLeuLeuSerPro 384  
 Db 1381 ATTATCAAGGCTACTAGCTCTCTCTCGGCAATGCGAGATGATGATTCATGCTCC 1440  
 QY 385 PheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTyrSerLeuPheThr 404  
 Db 1441 CTCTACTCCCTCTCTATATGCTCAGCTCTTGGCAGCAAGATCTTGTCTATGCTACTC 1500  
 QY 405 IleArgAsnAlaAspTrpGlyThr-----ArgLysLysLeuLeu 417  
 Db 1501 ATCAACAAGCTGTGGCTGGGCACTTCTGGCAGGAACCACTTGTGTC 1545  
 RESULT 2  
 LOCUS AK079729 2891 bp mRNA linear HTC 05-DEC-2002  
 DEFINITION Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:A430041J01 product:hyaluronan synthase 2, full insert sequence.  
 ACCESSION AK079729 1 GI:26348186  
 VERSION 1  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE 1  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE 2  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20493374  
 PUBMED 11042159  
 REFERENCE 3  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913  
 PUBMED 11076861  
 REFERENCE 4  
 AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schiraldi, L., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamya, M., Lee, N. H., Lyons, P., Marchionni, L., Washima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y.  
 TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL

21085660

11217851

PUBMED

5

REFERENCE

5

AUTHORS

Group Phase I & II Team.

TITLE

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL

Nature 420, 563-573 (2002)

REFERENCE

6 (bases 1 to 2891)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirao, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Onosato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES

source

location/Qualifiers

1..2891

organism="Mus musculus"

/mol\_type="mRNA"

strain="C57BL/6J"

/db\_xref="FANTOM\_DB:A430041J01"

taxon="10090"

/clone\_lib="A430041J01"

tissue\_type="thymus"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

dev\_stage="0 day neonate"

547..2205

note="unnamed protein product; hyaluronan synthase 2 (MGDI) MGI:107821, GBINM\_008216, evidence: BLASTN, 99%, match=2847)

putative

codon\_start=1

/protein\_id="BAC37733.1"

db\_xref="GI:26348187"

/translation="MHCEFLCVLRIGTLFGVSLLLGTAAYIVGVQFTQDNYIF SFGLYKFLASHLIQSLFALEHMKMKSLPTIKLNKVALCIAAYQEDPYLRK LQSVKRLTPGVKVMYVDGNDDLLYIMDFSEVMGRKSAIYIKWNHEKPGPT EESHKSSOHTVQLVLSNKSIMQKGGKREVMYTAFLALGRSDYVQVCDSDTMD PASSVMKVLDEEDPMVGGVGGVQVLLNKYSWISLSSVRYWMAFNERRACQSYGC VQCLSGPLNYSNLSLHEFVDFNOEFNQCSFGDDRLNTRVLSLGYATKATARS KCLTEPIELRLNQQTRWSKYFRELNYAMFHKHLMWYEAIVTGFPPFLIA TQVLFGRKLVNLIILFLVQLVLFKISFASCLRNIVMFMSLYSVLYMSLLPA KMFATINKAGTSGRTKIVVNFILIPVSVNFTILGGVIFITYKESKKPFSEK QTVLIVGTLLIYACVWMLLYLVLLINKGRRKKGGQYDMLDV"

polyA\_signal

2862..2867

note="putative"

2891

polyA\_site

/note="putative"

BASE COUNT

818 a 588 c 637 g 848 t

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

1.06e-36

439.00

46.65%

28.94%

20.47%

11

2891

134

82

183

65

14

US-09-469-200D-2 (1-417) x AK079729 (1-2891)

QY

7

LeuileThrValValAlaPheSerIlePheTrpValLeuLeuIleTyrValAsnValTyr 26

Db

613

CTCTCGGAATCACAGCTGCTTATATTTGGCTACCGATTTATCCAAACAGATATATAC 672

QY

27

LeuPheGlyAlaLysGlySerLeuSerIleTyrGlyPheLeuLeuIleAlaTyrLeuLeu 46

Db

673

TACTTC-----TCATTGGACTGACGGTGCCTTTTATAGCTGCATCTCATC 720

QY

47

Val-LysMetSerLeuSerPhePhe-----TyrLysProPheLysGlyVar 61

Db

721

ATCCAAGGCTCTTGGCTTTTGGACACCGGAAATGAAGAAGTCTCTTGAACCCCG 780

QY

61

gAlaglyInTyrLys---ValAlaIleIleProSerTyrAsnGluAspAlaGluSe 80

Db

781

ATTAATATG-AACAACACGGTAGCACTCTGCATCGCTGCTACCAAGAGGACCTGACTA 839

QY

80

LeuLeuGluThrLeuLysSerValGlnGlnThrTyrProLeuAlaGluIleTyrVa 100

Db

840

CTTACGGAATGTTGCAATCTGTGAAGAGCTGACCTCCTGGGATTAAAGTCGTGAT 899

QY

100

lValAspAspGlySerAlaAsp-----GluThrGlyIlely 112

Db

900

GGTCATCGATGGGAACCTCAGACGACGACCTTACATGNTGGACATATTCACCAAGTAT 959

QY

108

-----GluThrGlyIlely 112

Db

960

GGCAGGGACAAATCGCCACGTACATCTGGAGAACAACTTTCATGAAAGGAGCCTGG 1019

QY

112

saGileGluTyrValArgAspThrGlyAsp-----LeuSerSerAs 127

Db

1020

TGAGACAGAGAGNGTCCCATTAAGAAGTTACACACATGTACCCACCATGTGCTGTCTAA 1079

QY

127

nValIleValHisArgSerGluLysAsnGlnGlyHisArgHisAlaGlnAlaTrpAlaPh 147

Db

1080

CAAAAGTATTCATCATGCAAAATGGGTGGAAGAGAGAAAGTCATGTACACAGCCTT 1139

QY

147

eGluArg-----SerAspAlaAspValPheLeuThrValAspSerAspThrTyrIlely 165

Db

1140

CAGACACTGGGGGAGAGCGCTGATATGTACAGGTGTGTGACTCAGATATCTATGCTGA 1199

QY

165

rProAspAlaLeuGluLeuLeuLysThrPhe---AsnAspProThrValPheAlaAla 184

Db

1200

CCCTGCCTCATCTCTGGAGATGCTGAGGTCTTAGAGGAGAGACCTATGTTGGAGGTGT 1259

QY

184

aThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIl 204

Db

1260

TGGAGGAGATGCTCCAGATTTAAACAAGTATGATTCCTGCTCCTCTCCTCAGCAGGT 1319

QY

204

eArgTyrAspAsnAlaPheGlyValGluArgAlaAlaGlnSerValThrGlyAsnIlele 224

Db

1320

GATACTGGATGGCTTTTAATAGAAAGGCGCTGCCAGCTTATTTTGGCTGTGTCCA 1379

QY

224

uValCysSerGlyProLeuSerValTyrArgGluValValProAsnIleAspAr 244

Db

1380

GTCCATTAACGGCTCTCTGGGATGTACAGAACTCCTCTGCTGCTGATTAATTTGTGAGA 1439

QY

244

gTyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspArgCysLeuTh 264

Db

1440

CTGGTACAATCAGGAATTCATGGTAAACCAATGCAATGCGTGTGGTACGACGACCTTAC 1499

QY

264

rAsnTyrAlaThrAspLeuGly---LysThrValTyrGlnSerThrAlaLysCysIleTh 283

```

Db 1500 CAACAGGCTGTGAGTCTGGCTATGCACTAAATACAGCGCTGGTCCAAAGTGCCTTAC 1559
QY 283 rAspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsnLysSerPh 303
Db 1560 TGAACCTCCATAGATATCTCAGATGGCTGGAACAGCAGACCGGTTGGCAAGTCCCTA 1619
QY 303 ePheArgGluSerIleIleSerValLysLysLysMetAsnAsnProPheValAlaLeuTr 323
Db 1620 CTTCCGAGAGTGGCTGTACAAATGCCATGGTGTTCACAGCAT-----CACTTGTG 1670
QY 323 pThrIleLeuGluValSerMet-----PheMetLeuLeuValTyrSerVa 338
Db 1671 GATGACCTPATAAGCTGTTTATCATTGATCTTTTCTTTCTTCTTCCATTCGCCACAGTCA 1730
QY 338 lValAspPhePheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLe 358
Db 1731 CCAGCTCTTCTACAGGGGTAAATC-----TGG-----AACATCTCTCTCTCTCT 1775
QY 358 uValIleIlePheLeValAlaLeuCysArg---AsnIleHisTyrMetLeuLysHisPr 377
Db 1776 GTTAACCTGCCAGTAGTGGTCTCATCAAGTCATCTTTTGGCCAGCTGCCTTAGAGGAAA 1835
QY 377 oLeuSerPheLeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLe 397
Db 1836 TATCGTCATGATGATTCATCTCTCTATTCAGTGTATATACATGTCAAAGTCTACTTCTCTG 1895
QY 397 uLysLeuTyrSerLeuPheThrIleArgAsnAlaAspTrpGlyThr-----ArgLysLy 415
Db 1896 CRAAGTGTTCATTCACATTAACCAACAAAGCTGGGTGGGCACATCTGGAAGGAGAC 1955
QY 415 sLeuLeu 417
Db 1956 CATTTGTT 1962

RESULT 3
CA324064
LOCUS
DEFINITION
  UI-M-FY0-cco-g-16-0-01.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
  IMAGE: 6822281 5', mRNA sequence.
ACCESSION
  CA324064
VERSION
  EST.
KEYWORDS
  Mus musculus (house mouse)
SOURCE
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 809)
  NIH-MGC http://mgi.nci.nih.gov/
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  Tissue procurement: Dr. Jim Lin, University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LIN at:
  http://image.llnl.gov
  This clone was contributed by the Brain Molecular Anatomy Project
  (BMAP)
Seq primer: pYX-5.
Location/Qualifiers
  1..809
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strain="C57BL/6"
  /db_xref="taxon:10090"
  /clone IMAGE: 6822281
  /tissue_type="whole brain"
  /dev_stage="embryo 13.5, 14.5, 16.5, 17.5 dpc"
  /lab_host="DH10B (T1 phage resistant)"
  /clone_lib="NIH_BMAP_FY0"

```

## FEATURES

source

```

/note="Organ: Brain; Vector: pYX-Asc; Site.1: Ecor I;
Site.2: Not 1; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is ACGGACAG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
program coordinator."
BASE COUNT 175 a 218 c 193 g 221 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 1,94e-24 Length: 809
Score: 320.00 Matches: 76
Percent Similarity: 50.57% Conservative: 57
Best Local Similarity: 28.90% Mismatches: 96
Query Match: 14.92% Indels: 34
DB: 14 Gaps: 8

US-09-469-200D-2 (1-417) x CA324064 (1-809)
QY 166 ProAspAlaLeuGluLeuLeuLysThrPhe---AsnAspProThrValPheAlaAla 184
Db 12 CCAGCTGCACCATTTGAGATGCTTCGAGTCTTGGAGAAGATCCCAAGTAGAGGTGT 71
QY 185 ThrGlyHisLeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIle 204
Db 72 GGAGGAGATGTCCTCAATCTCCACAAAGATGATTCATGATCTCTCTCTGAGCAGGTG 131
QY 205 ArgTyrAspAlaPheGlyValGluArgAlaGlnSerValThrGlyAsnIleLeu 224
Db 132 AGGTACTGATGGCTTCAACGTGGAGCGGCTCCAGCTCTCTCTCTCTCTCTCTCTCT 191
QY 225 ValCysSerGlyProLeuSerValTyrArgArgGluValValProAsnIleAspArg 244
Db 192 TGTATTAGTGGCCCTTTGGCATGTACCGCAACAGCTCTCTCTCTCTCTCTCTCTCT 251
QY 245 TyrIleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspArgCysLeuThr 264
Db 252 TGGTACCATTACAGAGTCTCTAGGAGCAGATGCAGCTTTGGGATGATCGGACCTTACC 311
QY 265 AsnTyrAlaThrAspLeuGly---LysThrValTyrGlnSerThrAlaLysCysIleThr 283
Db 312 AACCGAGTCTCTGAGTCTTGGTACCGGACTAAGTATATACAGCAGCTCTAAGTGCCTC 371
QY 284 AspValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrpAsnLysSerPhe 303
Db 372 GAGACCCCACTAGTACTCTTCGATGGCTCATCATCAGAAACCCGCTGGAGCAGCTTAC 431
QY 304 PheArgGluSerIle-----IleSerVal 311
Db 432 TTTCCGAATGGCTCTACAAATTCCTGTGTGTTCCATAGCACCACCTCTGGATGACCTAT 491
QY 312 LysLysIleMetAsnAsn-----ProPheValAlaLeuThrPheLeuGluValSer 329
Db 492 GAATCAGTGGTCACAGGTTTCTTCCCATCTTCTCTCTCTCTCTCTCTCTCTCTCT 548
QY 330 MetPheMetLeuValTyrSerValValAspPheValGlyAsnValArgGluPhe 349
Db 549 -----TTCACCGTGGCGCATC----- 566
QY 350 AspTrpLeuArgValLeuAlaPheLeuValIleIlePheIleValAlaLeuCysArgAsn 369
Db 567 ---TGG---AACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 620
QY 370 IleHis---TyrMetLeuLysHisProLeuSerPheLeuLeuSerProPheTyrGlyVal 388

```

```

Db 621 ACCTATGCGCTGCTTCTTCGAGGCAATGCAGAGATGATCTTCATGTCNCCTCTACTCCCTT 680
    |||
Qy 389 LeuHisLeuPheValLeuGlnProLeuLysLeuThrSerLeuPheThrIleArgAsnAla 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 681 CTCATATGTCAGCGCTTTCGCGCAAGATCTTCTATTGCTTACCATCAACAAGTCT 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 409 AspTrpGly 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 741 GGCTGGGA 749

RESULT 4
BZ606767
LOCUS WHADE10TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
DEFINITION sapiens genomic clone MCF7_1-20A19, genomic survey sequence.
ACCESSION BZ606767
VERSION BZ606767.1 GI:31515328
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 980)
Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
Gray,J.W. and Collins,C.
ESP: a sequence-based approach to the structural genomics of tumors
JOURNAL Unpublished (2002)
COMMENT Contact: Volik SV
Colin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolik@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com
Class: BAC ends.
FEATURES
    source
        1..980
            Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /clone="MCF7_1-20A19"
                /sex="female"
                /clone_lib="Human MCF7 breast cancer cell line library
                (MCF7_1)"
                /note="Vector: pCRAC1; Site_1: HindIII; This library was
                constructed from MCF7 breast cancer cell line by Amplicon
                Express (http://www.genomex.com) using their standard
                procedure."
BASE COUNT 257 a 198 c 215 g 310 t
ORIGIN
Alignment Scores:
Pred. No.: 1.33e-23 Length: 980
Score: 313.50 Matches: 76
Percent Similarity: 52.72% Conservative: 50
Best Local Similarity: 31.80% Mismatches: 96
Query Match: 14.62% Indels: 17
DB: 29 Gaps: 7

US-09-469-200D-2 (1-417) x BZ606767 (1-980)
Qy 188 LeuAsnValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIleArgTrpAsp 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29 TTGCAGATTTTAAACAAGTACCATCTCTGGATCTTCATCTCCACAGAGTGAATATGG 88
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 208 AsnAlaPheGlyValGluArgAlaGlnSerValThrGlyAsnIleLeuValCysSer 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 ATGGCTTTTAAATATGAAGGCGCTGTCTTATTTTGGTGTCTTCTCAGTGCATTAGT 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Qy 228 GlyProLeuSerValTrpArgGluValValProAsnIleAspArgTrpIleAsn 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 GGACCTCTGGGATGATACAGAACTCTTGTTCATGAGTTTGGGAATGGTACAAAT 208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 248 GlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThrAsnTrpAla 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 CAAGAATTTATGGCAACCAATGATGTTGGTGTGACAGGAGCATCTCAGAACCGGGTG 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 268 ThrAspLeuGly--LysThrValTrpGlnSerThrAlaLysCysIleThrAspValPro 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 CTGAGCCTGGGCTATGCAACAATAACACAGCTCGATTAAGTCCCTTACCTGAAACACT 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 287 AspLysMetSerThrTrpLeuLysGlnAsnArgTrpAsnLysSerPhePheArgGlu 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 ATAGATATCTCAGATGGGTAAACACAGACACCCGTTGGAGCAAGTCTTACTTCCGAGA 388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 307 SerIleLeuSerValLysLysIleMetAsnAsnProPheValAlaLeuThrIleLeu 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 389 TGGCTGTACATGCAATGTGGTTTCACAAACAT-----CACTTGTGGATGACCTAC 439
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 327 GluValSerMet-----PheMetMetLeuValTrpSerValValAspPhe 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 440 GAAGCATATCTACTGATCTTCTCTCTTCTTCATTCGCCAGCAATGTCAGCTCTTC 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 342 PheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuValIle 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 500 TACCGGGGTAAAT-----TGG---AACATTCCTCTCTTCTTGTAACTGTC 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 362 PheIleValAlaLeuCysArg---AsnIleHisTrpMetLeuLysHisProLeuSerPhe 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 CAGCTAGTAGTCTCATATAAAATCATCTTTGCCAGCTGCCTTAGAGAAATATCGCTCATG 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 381 LeuLeuSerProPheTrpGlyValLeuHisLeuPheValLeuGlnProLeuLysLeuTrp 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 GCTTCATGCTCTCTACTACTGCTGTTATACATGCTCAGTTACTTCCGCCAGATGTTT 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 401 SerLeuPheThrIleArgAsnAlaAspTrpGlyThr-----ArgLysLysLeuLeu 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 GCAATTGCACATAAACAAGCTGGTGGGCGACATCAGGAAGGAACCACTGTT 721
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
BZ278152
LOCUS 603865094F1 CSEQCHN54 Gallus gallus cdna clone CHEST882124 5', mRNA
DEFINITION sequence.
ACCESSION BZ278152
VERSION BZ278152.1 GI:25727606
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Fong,W.T., Fickie,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1..711
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
FEATURES
    source

```



Db 99 CAGGTCTGTGTCGACACCAAGGTTGGACCCCGGAGCTGCTGGAGCTGCTGGCGGTA 158  
 QY 176 PheAsn---AspProThrValPheAlaAlaThrGlyHisLeuAsnValArgAsnArgGln 194  
 Db 159 CTGGACGAGGACCCCGGTAGGCTGTTGGTGGGAGCGTGGGATPCTTAACCTCTG 218  
 QY 195 ThrAsnLeuLeuThrArgLeuThrAspLeuThrAspAsnAlaPheGlyValGluArg 214  
 Db 219 GACTCTGGGTGAGCTTCTAGCAGCTCGATAGCTGGTGGTGGTGGTGGTGGTGG 278  
 QY 215 AlaAlaGlnSerValThrGlyAsnLeuValCysSerGlyProLeuSerValThrArg 234  
 Db 279 GCTTGTGAGAGTCTTCCACTGTGTATCTGTCATCAGCGTCTCTAGGCTATATAGG 338  
 QY 235 ArgGluValValProAsnLeuAspArgTyrIleAsnGlnThrPheLeuGlyIlePro 254  
 Db 339 ATATACCTTCTGAGCAGTTCTTACAGCCTGGTACACCAAGTTCTTGGGTATCCAC 398  
 QY 255 ValSerIleGlyAspAspArgCysLeuThrAsnTyrAlaThrAspLeuGly---LysThr 273  
 Db 399 TGTACTTTTGGGATGACCGGCACTCACCACCGCATGCTCAGCATGGGTATGCTACC 458  
 QY 274 ValTyrGlnSerThrAlaLysCysIleThrAspValProAspLysMetSerThrTyrLeu 293  
 Db 459 AAGTACACTCCAGGTCCTGCTGCTACTCAGAGCGGCTGCTCTTCTGCGGTGGCTG 518  
 QY 294 LysGlnGlnAsnArgTyrAsnLysSerPhePheArgGluSerIle----- 308  
 Db 519 ACCGAGCAGACGCTGGTCCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 578  
 QY 309 -----IleSerValLysLysIleMetAsnAsn-----ProPhe 319  
 Db 579 TGGCACCAGGACCATCGCTGGATACCTACGAGGCGGTGCTCCGCCCTGTTCCTCCTC 638  
 QY 320 ValAlaLeuThrPheIleLeuValSerMetPheMetMetLeuValTyrSerValVal 339  
 Db 639 TTCGTGGCGCCACTGTGCTGCTGCTG----- 665  
 QY 340 AspPhePheValGlyAsnValArgGluPheAspTyrLeuArgValLeuAlaPheLeuVal 359  
 Db 666 ---TTCATACGGCGCGCCCTTGGCGCTGCTGGGTGCTGCTGCTGCTGCTGCTGCTG 722  
 QY 360 IleLeuPheLeuValAlaLeuCysArgAsnIleHisTyrMetLeuLysHisProLeuSer 379  
 Db 723 GCACCTGGCAGGCGCCCTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 782  
 QY 380 PheLeuLeuSerProPheTyr-----GlyValLeuHisLeuPheValLeuGlnProLeu 397  
 Db 783 TGTGTTCTACGGCCCTTCTACATGCGGGGCTCTCTG-----CCTGCC 824  
 QY 398 LysLeuTyrSerLeuPheThrIleArgAsnAlaAspTyrGlyThr-----ArgLysLys 415  
 Db 825 AGTTCTGGCGCTATCCCTGTAACCGAGCGGTGGGCGCTTCCGCGCGCGGGA 884  
 QY 416 Leu 416  
 Db 885 CTG 887

RESULT 7

LOCUS BX371636  
 DEFINITION BX371636 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 clone CS0D1027J02 3-PRIME, mRNA sequence.  
 ACCESSION BX371636  
 VERSION BX371636.1  
 KEYWORDS GI:30450029  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 915)  
 AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
 TITLE Full-length cDNA libraries and normalization

JOURNAL COMMENT

Unpublished  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 9502.r. For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0BA10192C12\_CS01819\_1&cluster=9502.r.  
 Contact : Feng liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Paraday Avenue Genoscope sequence ID : CS0BA10192C12\_CS01819\_1.

FEATURES

source

Location/Qualifiers  
 1..915  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1027J02"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 181 a 302 c 268 g 155 t 9 others

ORIGIN

Alignment Scores:  
 Pred. No.: 1.26e-20 Length: 915  
 Score: 286.00 Matches: 73  
 Percent Similarity: 50.70% Conservative: 36  
 Best Local Similarity: 33.95% Mismatches: 86  
 Query Match: 13.33% Indels: 20  
 DB: 13 Gaps: 5

US-09-469-200D-2 (1-417) x BX371636 (1-915)

QY 119 ArgAspThrGlyAspLeuSerSerAsnValIleValHisArg----- 132  
 Db 686 AAGATCTCTGGCGCGCTGGCAGTGGAGCGCTGTTGAGGACTCGCAGGTGCTGCTGCTG 627  
 QY 133 SerGlyLysAsnGlnGlyLysArgHisAlaGlnAlaThrAlaPheGluArg-----Ser 150  
 Db 626 GCACGCGCTGGCGCGCAACGCGCGGCTCATGTACACAGCTTCAGGCGCTCGGAGAT 567  
 QY 151 AspAlaAspValPheLeuThrValAspSerAspThrTyrIleTyrProAspAlaLeuGlu 170  
 Db 566 TCAGTGGACTACGTGCAGGTCTGTGACTCGCACACAAGGTGGACCCCATGCTGCTGCTG 507  
 QY 171 GluLeuLeuLysThrPheAsn---AspProThrValPheAlaAlaThrGlyHisLeuAsn 189  
 Db 506 GAGCTCTGGCGGTACTGGACGAGGACCCCGGTAGGGGCTGTTGGTGGGAGCGTGGCG 447  
 QY 190 ValArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIleArgTyrAspAsnAla 209  
 Db 446 ATCTTAACTCTCTGGACTCTCTGGGTGAGTCTTCTAAGCAGCTTCGCTACTGGGTAGCC 387  
 QY 210 PheGlyValGluArgAlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGlyPro 229  
 Db 386 TTCAATGTGGAGCGGCTTGTGACAGCTACTTCCACTGTGTATCTGTCATCAGCGTCT 327  
 QY 230 LeuSerValTyrArgGluGluValValProAsnIleAspArgTyrIleAsnGlnThr 249  
 Db 326 CTAGGCTATATAGGAATAACTCTTTCGACGAGTTCCTTGGAGGCTGTGTACACCAAG 267  
 QY 250 PheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThrAsnTyrAlaThrAsp 269  
 Db 266 TTCTCTGGGTACCTACTCTACTTTTGGGATGACGCGCAGCTTCACACCCGCTGCTCAGC 207  
 QY 270 LeuGly---LysThrValTyrGlnSerThrAlaLysCysIleThrAspValProAspLys 288  
 Db 206 ATGGGTTATGTACCAAGTACACCTCCAGGTCCCGCTGCTGCTACTCAGAGCGCCCTCGTCC 147

QY 289 MetSerThrTyrLeuLysGlnGlnAsnArgTrrPAsnLysSerPhePheArgGluSerlle 308  
 Db 146 TTCTCGGTGGTGGAGCCAGACAGCTGGTCCAGTTCCTCCGTGAGTGGCTG 87  
 QY 309 IleSerValLysLysIleMetAsnAsnProPheValAlaLeuTrrP 323  
 Db 86 -----TACACGGCTCTGG 72

RESULT 8  
 AL775141  
 LOCUS AL775141 639 bp mRNA linear EST 25-JUN-2002  
 DEFINITION XGC-gastrula silurana tropicalis cDNA clone TGas077a10 5', mRNA sequence.  
 ACCESSION AL775141  
 VERSION AL775141.1 GI:21560845  
 KEYWORDS EST.  
 SOURCE Silurana tropicalis (western clawed frog)  
 ORGANISM Silurana tropicalis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Silurana.  
 REFERENCE 1 (bases 1 to 639)  
 AUTHORS Taylor, R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.  
 TITLE Sanger Xenopus tropicalis EST project 2002  
 JOURNAL Unpublished  
 COMMENT Contact: Taylor R  
 Sanger Centre  
 Hinxton, Cambridgeshire, CB10 1SA, UK  
 Email: trop@sanger.ac.uk  
 Sanger Xenopus tropicalis EST project 2001  
 TROPICALIS\_SEQUENCE\_ID: TGas077a10.plksp6  
 Sequencing primer: plksp6  
 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
 FEATURES  
 source  
 1..639  
 Location/Qualifiers  
 /organism="Silurana tropicalis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8364"  
 /clone="TGas077a10"  
 /dev\_stage="gastrula (stages 10.5-13 mixed)"  
 /lab\_host="Escherichia coli XL1-blue"  
 /clone\_lib="XGC-gastrula"  
 /note="vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dT primed from 5' of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."  
 BASE COUNT 160 a 152 c 173 g 154 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,24e-20 Length: 639  
 Score: 284.00 Matches: 68  
 Percent Similarity: 52.06% Conservative: 33  
 Best local Similarity: 35.05% Mismatches: 79  
 Query Match: 13.24% Indels: 14  
 DB: 9 Gaps: 4

US-09-469-200D-2 (1-417) x AL775141 (1-639)

QY 134 GluLysAsnGlnGlyLysArgHisAlaGlnAlaTrrPAlaPheGluArg-----SerAsp 151  
 Db 13 CAAAATGGGGGAGAGAGAGATCATGTACACAGCTTCAAGCAGCTGGGGAAGC 72  
 QY 152 AlaAspValPheLeuThrValAspSerAspThrTyrIleTrrProAspAlaLeuGlu 171  
 Db 73 GTGGATATGTGCAGGTATGCGACTCTGCACAGTGTGCGCATCGCGTGGAG 132  
 QY 172 LeuLeuLysThrPhe---AsnAspProThrValPheAlaAlaThrGlyHisLeuAnVal 190  
 Db 133 ATGCTCAAGTACTGGAGGAGACATCATGTTGGAGGAGTGGGTGGAGATGTGCAGATT 192

QY 191 ArgAsnArgGlnThrAsnLeuLeuThrArgLeuThrAspIleArgTrrAspAsnAlaPhe 210  
 Db 193 TTAACAAGTACGACTCGTGGATTCTCTCTGAGTAGCGTCAGATACGATGGGCTTT 252  
 QY 211 GlyValGluArgAlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGlyProLeu 230  
 Db 253 AACATTGAGAGCGGTGGCGTCTTACTTGGCGTGTGCAATGCATCAGCGGCCCTTG 312  
 QY 231 SerValTrrArgGluValValProAsnIleAspArgTrrIleAsnGlnThrPhe 250  
 Db 313 GGGATGACGGAGCTCTCTCCAGCAATCTCTGAGATGGTGCACCAAGATTT 372  
 QY 251 LeuGlyIleProValSerIleGlyAspArgCysLeuThrAsnTrrAlaThrAspLeu 270  
 Db 373 TTGGGTTCCCGAGTGCAGTTTCGGGGATGATCGCCACCTCACCACCGAGTCTTGAGTCTG 432  
 QY 271 Gly---LysThrValTrrGlnSerThrAlaLysCysIleThrAspValProAspLysMet 289  
 Db 433 GGCTACGCAACCAATACACGGCCAGTCCAAATGCTTACTGAAACCCCAACCGAGTAC 492  
 QY 290 SerThrTrrLeuLysGlnGlnAsnArgTrrPAsnLysSerPhePheArgGluSerIlelle 309  
 Db 493 CTACGATGGCTCAACCAACGACGCGGTGGAGCAAGTCTTCTCCGAGATGGCTG--- 549  
 QY 310 SerValLysLysIleMetAsnAsnPropheValAlaLeuTrrP 323  
 Db 550 -----TACAATGCAATGTGG 564

RESULT 9  
 BG707429  
 LOCUS BG707429 551 bp mRNA linear EST 07-MAY-2001  
 DEFINITION 602672848f1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:4795643 5', mRNA sequence.  
 ACCESSION BG707429  
 VERSION BG707429.1 GI:13983769  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
 1 (bases 1 to 551)  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: csapbs@remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshituki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM10678 row: p column: 12  
 High quality sequence stop: 551.  
 Location/Qualifiers  
 1..551  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4795643"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_96"  
 /notes="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtgag ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library



```

QY 294 LysGlnGlnAsnArgTrrpAsnLysSerPhePheArgGluSerIleIleSerValLysLys 313
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 491 ACCCAGCAGACGCTGGGCCAGTCTTCCGAGTGGCTG----- 535
QY 314 IleMetAsnAsnPropheValAlaLeuTrp 323
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 -----TACAAGCGCTCTGG 550

RESULT 11
BJ041963 656 bp mRNA linear EST 07-DEC-2001
LOCUS
DEFINITION
  BJ041963 NIBB Mochii normalized Xenopus neurula library Xenopus
  laevis cDNA clone XL050c20 5', mRNA sequence.
ACCESSION
  BJ041963
VERSION
  BJ041963.1 GI:17417509
KEYWORDS
  EST.
SOURCE
  Xenopus laevis (African clawed frog)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
  Xenopodinae; Xenopus.
REFERENCE
  1 (bases 1 to 656)
AUTHORS
  Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i.T. and Kohara
  ,Y.
TITLE
  Expressed genes in X. laevis embryo
JOURNAL
  Unpublished
COMMENT
  Contact: Tadasu Shin-i
  Center For Genetic Resource Information
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshini@genes.nig.ac.jp.
FEATURES
  source
  1..656
  /organism="Xenopus laevis"
  /mol_type="mRNA"
  /db_xref="taxon:8355"
  /clone="XL050c20"
  /tissue_type="whole embryo"
  /dev_stage="stage 15"
  /clone_lib="NIBB Mochii normalized Xenopus neurula
  library"
BASE COUNT 159 a 169 c 142 g 184 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 1.83e-19 Length: 656
Score: 273.50 Matches: 74
Percent Similarity: 51.40% Conservative: 36
Best Local Similarity: 34.58% Mismatches: 89
Query Match: 12.75% Indels: 16
DB: 12 Gaps: 6

US-09-469-200d-2 (1-417) x BJ041963 (1-656)
QY 206 TyrAspAsnAlaPheGlyValGluArgAlaGlnSerValThrGlyAsnIleLeuVal 225
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 33 TACTGGATGGGCTTAACTAGAGAGCATGCCAGTCTTACTTGGCTGTGCAATGC 92
QY 226 CysSerGlyProLeuSerValTyrArgGluValValProAsnIleAspArgTyr 245
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 93 ATTACGCCCGCTGGGGGTACCGCAATTCCTCTCCACGAATTCATGAAGACTGG 152
QY 246 IleAsnGlnThrPheLeuGlyIleProValSerIleGlyAspAspArgCysLeuThrAsn 265
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153 TACACCAAGAAATTTGGTTCCTCCAGTGCAGTTTGGGGAGACCTCCACTACCAAT 212
QY 266 TyrAlaThrAspLeuGly---LysThrValTyrGlnSerThrAlaLysCysIleThrAsp 284
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 213 CGAGTTTGTAGCTCTGGCTATGCCAACCAATACACGGCCAGATCCCAAAATGCTTACTGAA 272
QY 285 ValProAspLysMetSerThrTyrLeuLysGlnGlnAsnArgTrrpAsnLysSerPhePhe 304

```

```

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 ACACCACCGAGTACCTACGCTGGCTCAACAGCAACGCGATGGAGCAAGTCTCTTC 332
QY 305 ArgGluSerIleIleSerValLysLysIleMetAsnAsnPropheValAlaLeuTrpThr 324
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 333 CGAGATGGCTGTACAATTCATTGTGGTTCCATAACAT-----CACTTATGGATG 383
QY 325 IleLeuGluValSerMet-----PheMetMetLeuValTyrSerValVal 339
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 ACCTAGCAAGCTGTGATTACTGGATTCTTCTCTCTCATCGCCACTGTCCACG 443
QY 340 AspPheValGlyAsnValArgGluPheAspTrpLeuArgValLeuAlaPheLeuVal 359
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 CTCTTCTACCGTGGAGATC-----TGG---AACATCTCTCTGTCTTGTGCTG 488
QY 360 IleIlePheIleValAlaLeuCysArg---AsnIleHisTyrMetLeuLysHisProLeu 378
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 489 ACAGTACAACCTGTAGGCTTATCAAACTCTCTCTAGTGCCTCCGAGGGAACATA 548
QY 379 SerPheLeuLeuSerProPheTyrGlyValLeuHisLeuPheValLeuGlnProLeuLys 398
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 549 GTCATGGTCTTCATGCTCTCTTACTACAGTGTATACATGTCACGTTTACTACCTGCCAAA 608
QY 399 LeuTyrSerLeuPheThrIleArgAsnAlaAspTrrpGlyThr 412
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 609 ATGTTGCCATTGCC-ACCATCACAGCAGGTTGGGCACA 649

RESULT 12
BJ062498 536 bp mRNA linear EST 07-DEC-2001
LOCUS
DEFINITION
  BJ062498 NIBB Mochii normalized Xenopus tailbud library Xenopus
  laevis cDNA clone XL068c24 5', mRNA sequence.
ACCESSION
  BJ062498
VERSION
  BJ062498.1 GI:17424132
KEYWORDS
  EST.
SOURCE
  Xenopus laevis (African clawed frog)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
  Xenopodinae; Xenopus.
REFERENCE
  1 (bases 1 to 536)
AUTHORS
  Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i.T. and Kohara
  ,Y.
TITLE
  Expressed genes in X. laevis embryo
JOURNAL
  Unpublished
COMMENT
  Contact: Tadasu Shin-i
  Center For Genetic Resource Information
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshini@genes.nig.ac.jp.
FEATURES
  source
  1..536
  /organism="Xenopus laevis"
  /mol_type="mRNA"
  /db_xref="taxon:8355"
  /clone="XL068c24"
  /tissue_type="whole embryo"
  /dev_stage="stage 25"
  /clone_lib="NIBB Mochii normalized Xenopus tailbud
  library"
BASE COUNT 137 a 120 c 134 g 144 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1.55e-19 Length: 536
Score: 273.00 Matches: 60
Percent Similarity: 59.00% Conservative: 27
Best Local Similarity: 40.00% Mismatches: 61
Query Match: 12.73% Indels: 2
DB: 12 Gaps: 2

```



## ACCESSION

CA988122  
CA988122.1 GI:27520797

## KEYWORDS

EST.

## SOURCE

Xenopus laevis (African clawed frog)

## ORGANISM

Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus.

## REFERENCE

1. (bases 1 to 866)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

## AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

## TITLE

Tumor gene index

## JOURNAL

Unpublished

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LRAM14486 row: k column: 02  
High quality sequence stop: 744.  
Location/Qualifiers

## FEATURES

source

1. 866  
/organism="Xenopus laevis"

/mol\_type="mRNA"

/db\_xref="taxon:8355"

/clone="IMAGE:6864987"

/tissue\_type="embryo (stage 10)"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NICHD\_XGC\_Emb1"

/note="vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI;  
Cloned unidirectionally. Primer: Oligo dt. Average insert  
size 1.55 kb. Constructed by Life Technologies. Note: This  
is a Xenopus Gene Collection (XGC) library."

## BASE COUNT

204 a 215 c 213 g 234 t

## ORIGIN

## Alignment Scores:

Pred. No.: 3.91e-18 Length: 866  
Score: 263.00 Matches: 76  
Percent Similarity: 45.79% Conservative: 49  
Best local Similarity: 27.84% Mismatches: 112  
Query Match: 12.26% Indels: 36  
DB: 14 Gaps: 8

US-09-469-200D-2 (1-417) x CA988122 (1-866)

QY 138 GlyLysArgHisAlaGlnAlaThrPalaPheGluArg-----SerAspAlaAspValPhe 155  
|||||  
Db 63 GGAAGAGAGAGTCTATGATACAGCATTCAGGCCATTCGGGACTTCGTGGACTATGTA 122  
QY 156 LeuThrValAspSerAspThrTyrIleTyrProAspAlaLeuGluLeuLeuLysThr 175  
|||||  
Db 123 CAGCTCTGTACTCGGACCAAACTGATGATGACTGCACTGCACTGCACTGCACTGCACT 182  
QY 176 PheAsnAspProThrValPheAlaAlaThr---GlyHisLeuAsnValArgAsnArgGln 194  
|||||  
Db 183 CTGGAGTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 242  
QY 195 ThrAsnLeuLeuThrArgLeuThrAspIleArgTyrAspAsnAlaPheGlyValGluArg 214  
|||||  
Db 243 GATTCCTTCATTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 302  
QY 215 AlaAlaGlnSerValThrGlyAsnIleLeuValCysSerGlyProLeuSerValTyrArg 234  
|||||  
Db 303 GCTGCGGAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362  
QY 235 ArgGluValValProAsnIleAspArgTyrIleAsnGlnThrPheLeuGlyIlePro 254  
|||||  
Db 363 AACACATCTCCAGGTGTTTGGAGCCTGTTACAGACAGAAATTTTGGACCTAT 422

QY 255 ValSerIleGlyAspAspArgCysLeuThrAspTyrAlaThrAspLeuGly---LysThr 273  
|||||  
Db 423 TGACTTTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 482  
QY 274 ValTyrGlnSerThrAlaLysCysIleThrAspValProAspLysMetSerThrTyrLeu 293  
|||||  
Db 483 AAATACACCCACAAATCCAGAGCATTCGGAATCCATCCCTGTATCTCCGCTGGTTG 542  
QY 294 LysGlnGlnAsnArgTyrPasnLysSerPhePheArgGluSerIleIleSerValLysLys 313  
|||||  
Db 543 AACACGAAACCCGCTGGACCAAGTCTTCTTCTGAGAGTGGCTGTATATGCCCAGTGG 602  
QY 314 IleMetAsnAspProPheValAlaLeuThrPheLeuGluValSerMetPheMetMet 333  
|||||  
Db 603 TGGCACAACGAT-----CACATCGATGACCTATGAG----- 635  
QY 334 LeuValTyrSerValValAspPhe----- 341  
|||||  
Db 636 -----TCTGTGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 686  
QY 342 -----PheValGlyAsnValArgGluPheAspThrLeuArgValLeuAlaPheLeuVal 359  
|||||  
Db 687 CTCATCTATGCGGACCATATGGAATGTGTGGCTC-----CTCCG 731  
QY 360 IleIlePheIleValAlaLeuCysArgAsnIleHis---TyrMetLeuLysHisProLeu 378  
|||||  
Db 732 TGCATCCAGATCATCTCTCTTCAATCCATCATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 791  
QY 379 SerPheLeuLeuSerPhePheTyrGlyValLeuLeuHisLeu 391  
|||||  
Db 792 ATTATGCTCTGATGCT 830  
RESULT 15  
LOCUS BU907404  
DEFINITION AGENCOURT\_10482681 NICHD\_XGC\_Emb1 Xenopus laevis cDNA clone  
IMAGE:6632567 5', mRNA sequence.  
ACCESSION BU907404  
VERSION BU907404.1 GI:24089318  
KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus.  
REFERENCE 1 (bases 1 to 876)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LRAM14202 row: b column: 23  
High quality sequence stop: 614.  
Location/Qualifiers  
1. 876  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="IMAGE:6632567"  
/tissue\_type="embryo (stage 10)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NICHD\_XGC\_Emb1"  
/note="vector: pCMV-SPORT6; Site.1: NotI; Site.2: SalI;  
Cloned unidirectionally. Primer: Oligo dt. Average insert  
size 1.55 kb. Constructed by Life Technologies. Note: This  
is a Xenopus Gene Collection (XGC) library."

Search completed: October 2, 2003, 16:10:25  
Job time : 2829 secs